Shane R. Jimerson
Matthew K. Burns
Amanda M. VanDerHeyden
Editors

Handbook of Response to Intervention

The Science and Practice of Multi-Tiered Systems of Support

Second Edition

Foreword by Sharon Vaughn

Springer

Handbook of Response to Intervention

Shane R. Jimerson • Matthew K. Burns
Amanda M. VanDerHeyden
Editors

Handbook of Response to Intervention

The Science and Practice of Multi-Tiered Systems of Support

Second Edition

Springer

Editors
Shane R. Jimerson
University of California, Santa Barbara
Gevirtz Graduate School of Education
Santa Barbara
CA
USA

Amanda M. VanDerHeyden
Education Research and Consulting, Inc.
Fairhope
AL
USA

Matthew K. Burns
University of Missouri
Columbia
MO
USA

ISBN 978-1-4899-7567-6 ISBN 978-1-4899-7568-3 (eBook)
DOI 10.1007/978-1-4899-7568-3

Library of Congress Control Number: 2015936035

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

This second edition of the handbook is dedicated to the many professionals who work diligently to educate and enhance the success of students and also to the scholars who inform our understanding of promoting the social and cognitive competence of students at school. By bringing the best of science to professional practice and highlighting lessons learned from implementation efforts across the country, it is hoped that the information presented in this handbook serves as a catalyst that advances the science and practice of assessment and intervention at school and, ultimately, promotes enhanced student outcomes.

Foreword

This handbook of response to intervention (RTI) exceeds all of the lofty goals typically used to describe topical handbooks. Shane Jimerson, Matthew Burns, and Amanda M. VanDerHeyden have edited a book assembling an outstanding group of contributors that form the foundation and pillars for our knowledge on response to intervention or multi-tiered systems of support. Additionally they have organized the content of the book to cover every issue related to response to intervention including the scientific foundations of RTI, psychometric measurement related to RTI, the role of consultation, monitoring response to intervention, using technology to facilitate RTI, and RTI and transition planning. In particular, I am impressed with the emphasis on both problem-solving and standardized approaches to RTI—as well as the breadth of coverage of assessment, progress monitoring, and interventions. Unlike many handbooks, this one provides critical information addressing issues for a range of individuals including school leaders, school psychologists, social workers, counselors, academic specialists, general education, and special education teachers. Anyone working on understanding RTI, whether as a scholar conducting research on this topic or a school practitioner searching for solutions to problems related to successful implementation of RTI, will find research-based practice knowledge in this text.

RTI or multi-tiered systems of support provide a framework for screening students with academic and behavior problems across all grades but with an emphasis on identifying students early who require additional instruction. Within the RTI framework are research-based systems for providing intensive interventions to accelerate students' progress, assuring all students are provided with high-quality instructions they need to meet the challenging goals of postsecondary education. RTI is beneficial to all educators not just those who have high numbers of students at risk because it assures that students' educational and behavioral needs are monitored with an action-plan for improved outcomes. While there are many ways to implement RTI well, supporting implementation of RTI is essential because it provides a safety net for our most vulnerable students.

Busy professionals are bombarded with information and sifting through it to determine sources that are worth reading—even studying—is a challenging task. This book is a resource worth reading for even the busiest professional.

Sharon Vaughn
University of Texas

Acknowledgments

The editors would like to acknowledge the exceptional efforts of colleagues who contributed to this handbook as authors, and also to Judy Jones and Garth Haller at Springer who were instrumental in bringing this handbook to print. In addition, we are grateful for the extraordinary efforts of numerous individuals who provided reviews of chapters, which further enhanced the quality and contents. The collective efforts of all those involved have resulted in an incredibly informative and valuable second edition of the handbook.

Finally, we acknowledge the tremendous support, inspiration, and perspective that our families provided throughout the development of this handbook. The collective adventures, celebrations, and challenges we have shared have enriched each of us. Our sincere appreciation to the following contributors: Kathryn O'Brien, Gavin Jimerson, Taite Jimerson, Mary Beth Burns, Kate Burns, Matthew Burns, Jr., Chad VanDerHeyden, Benjamin VanDerHeyden, and Kate VanDerHeyden.

Contents

About the Editors

Shane R. Jimerson Ph.D. is the chair and professor in the Department of Counseling, Clinical, and School Psychology at the University of California, Santa Barbara. Currently, Dr. Jimerson is the editor of *School Psychology Quarterly* published by the American Psychological Association, and President-Elect of the International School Psychology Association. He was also recently the president *of Division 16 (School Psychology) of the American Psychological Association*. He has contributed over 200 presentations during the past 15 years to diverse audiences of students, scholars, and professionals across more than 25 countries, including, Australia, Bangladesh, Canada, China, Denmark, England, Estonia, Finland, France, Germany, Greece, Hong Kong, India, Ireland, Jamaica, Japan, Latvia, Lithuania, Macau, Malta, Puerto Rico, the Netherlands, New Zealand, Portugal, Spain, and the USA. Dr. Jimerson is the co-founder of the *International Institute of School Psychology* (http://mina.education.ucsb.edu/jimerson/IISP/index.html). Among over 300 publications, Dr. Jimerson is the editor of and contributor to a special issue of *The California School Psychologist* journal addressing the topic of response to intervention and problem-solving strategies, entitled, *Response to Interventions Approaches: Supporting Early and Sustained Success for all Students*. He is also the lead-editor of *The Handbook of School Violence and School Safety: International Research and Practice 2nd Edition* (2012, Routledge), co-editor of *Best Practices in School Crisis Prevention and Intervention 2nd Edition* (2012, National Association of School Psychologists), *The Handbook of Bullying in Schools: An International Perspective* (2010, Routledge), the lead-editor of *The Handbook of International School Psychology* (2007, SAGE Publishing), and the lead editor of *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (2007, Springer Science). He is also co-author of *School Crisis Prevention and Intervention: The PREPaRE Model* (2009, National Association of School Psychologists), a co-author of a five-book grief support group curriculum series *The Mourning Child Grief Support Group Curriculum* (2001, Taylor and Francis), co-author of *Identifying, Assessing, and Treating Autism at School* (2006, Springer Science), co-author of *Identifying, Assessing, and Treating Conduct Disorder at School* (2008, Springer Science), co-author of Identifying, Assessing, and Treating PTSD at School (2008, Springer Science), co-author of *Identifying, Assessing, and Treating ADHD at School* (2009, Springer Science), and co-author of the *Promoting Positive Peer Relationships (P3R): Bullying Prevention Program* (2008, Stories of Us). He has also served as the editor of *The California School Psychologist* journal, associate editor of *School Psychology Review*, and the editorial boards of numerous journals including the *Journal of School Psychology and School Psychology Quarterly*. Dr. Jimerson has chaired and served on numerous boards and advisory committees at the state, national, and international levels, including, vice president for *Convention Affairs and Public Relations of Division 16 (School Psychology) American Psychological Association*, chair of the *Research Committee* of the *International School Psychology Association*, chair of the *Division 16 (School Psychology) conference proceedings for the American Psychological Association* conference, and chair of the *School Psychology Research Collaboration Conference*. The quality and contributions of his scholarship are reflected in the numerous awards and recognition that he has received. Dr. Jimerson received the *Best Research Article of the Year Award* from the *Society for the Study of School Psychology*, in 1998 and then again in 2000. He also received the 2001 *Outstanding Article of the Year Award* from the National Association of School Psychologists', *School Psychology Review*. Also in 2001, he was elected to membership in the *Society for the Study of School Psychology*. Dr. Jimerson's scholarly efforts were also recognized by the *American Educational Research Association* with the 2002 *Early Career Award in Human Development*. He and his UCSB research team received the 2003 *Outstanding Research Award* from the *California Association of School Psychologists*. Also during 2003, Dr. Jimerson received the *Lightner Witmer Early Career Contributions Award* from *Division 16 (School Psychology) of the American Psychological Association*. He and his UCSB research team also received the 2004 *Outstanding Research Award* from the *California Association of School Psychologists*. In 2006, Dr. Jimerson received the *President's Award for Exemplary Contributions from the California Association of School Psychologists*. In 2007, Dr. Jimerson was elected fellow of the *American Psychological Association, Division 16 (School Psychology)*. Dr. Jimerson received the 2010 Outstanding Contributions Award from the *American Psychological Association, Division 16 (School Psychology)*. Dr. Jimerson presented the 2012 Evan Brown Distinguished Lecture at the University of Nebraska, Omaha and also received the 2012 Ronda Talley Distinguished Leader and Advocate Award from Indiana University. In 2013 Dr. Jimerson was elected Fellow of the *American Psychological Association, Division 52 (International Psychology)* and also received the Outstanding Contributions Award from the *American Psychological Association, Division 16 (School Psychology)*. Dr. Jimerson received the 2014 Award of Excellence for Distinguished Contributions to School Crisis Management from the *National Association of School Psychologists* and he also received the 2014 Outstanding International Psychologist Award from the *American Psychological Association, Division 52 (International Psychology)*. His international professional and scholarly activities aim to advance and promote science, practice, and policy relevant to school psychology, in an effort to benefit children, families, and communities across the country and throughout the world.

Matthew K. Burns Ph.D. is the Associate Dean for Research and Professor of School Psychology with the College of Education at the University of Missouri. Dr. Burns has published over 200 articles and book chapters in national publications, and has procured over $ 11 million of external funding to support his research. He has also co-authored or co-edited 12 books including *Advanced RTI Applications, Volumes 1 and 2, Response to Intervention Implementation in Elementary and Secondary Schools: Procedures to Assure Scientific-Based Practices* (2nd edition), *A Guide to Refining and Retooling School Psychological Practice in the Era of RtI, Single Case Design for Measuring Response to Educational Intervention*, and *Curriculum-Based Assessment for Instructional Design: Using Data to Individualize Instruction*. Dr. Burns is the editor of *School Psychology Review* and past editor of *Assessment for Effective Intervention*. He has received numerous awards including the 2013 Discovery Award for Distinction in Educational Research from the Midwest Instructional Leadership Council and the 2011 Evidence of Scientific Research Education Award from the Minnesota Consortium for Evidence in Education. Finally, Dr. Burns is a highly sought after national speaker and has delivered over 70 invited or keynote presentations to national, regional, or state conferences, and has worked with dozens of schools across the country in implementing response to intervention.

Amanda M. VanDerHeyden Ph.D. is a private consultant and researcher who has directed and evaluated numerous school-wide intervention and reform efforts and her work has been featured on "Education News Parents Can Use" on PBS and The Learning Channel. Dr. VanDerHeyden has held faculty positions at Louisiana State University Health Sciences Center and University of California at Santa Barbara. She is the president of Education Research & Consulting, Inc. in Fairhope. Dr. VanDerHeyden serves as a former scientific advisor to the RTI Action Network at the National Center for Learning Disabilities, iSTEEP (a web-based data management system), and the Center on Innovations in Learning. She is a former guest and standing panel member for the Institute for Education Sciences at the U.S. Department of Education, and serves on the board of trustees for the Southwest Development Laboratory (SEDL, one of 10 regional laboratories funded by the US Department of Education). Dr. VanDerHeyden has published more than 70 scholarly articles and chapters, 6 books, and has given keynote addresses to state school psychology associations and state departments of education in 22 states. She is the co-author of the Evidence-Based Mathematics Innovation Configuration for the National Comprehensive Center for Teacher Quality at Vanderbilt University and now the Collaborator for Effective Education Development, Accountability, and Reform at University of Florida. Her most recent book (*The RTI Approach to Evaluating Learning Disabilities*) was featured at a forum for policymakers hosted by the National Center for Learning Disabilities as a best-practice guide for identifying and serving children with Learning Disabilities in October of 2013 in New York, NY. She actively conducts research focused on improving learning outcomes for students and her scholarly work has been recognized in the form of article of the year award in 2007 from *Journal of School Psychology*, the Lightner Witmer Early Career

Contributions Award from Division 16 (School Psychology) of the American Psychological Association, and her 2012 induction into the 100-member Society for the Study of School Psychology.

Contributors

Jill Allor Southern Methodist University, Dallas, USA

David H. Allsopp Tampa, USA

Julie Alonzo University of Oregon, Eugene, USA

Scott P. Ardoin Department of Educational Psychology, University of Georgia, Athens, Georgia

Jane Atwater University of Kansas, Kansas City, KS, USA

Kathryn E. Bangs Department of Educational Psychology, University of Georgia, Athens, Georgia

Kerry A. Bollman TIES, Cleveland Heights, USA

Tracy A. Bradfield University of Minnesota, Minneapolis, MN, USA

Tera L. Bradley Boling Center for Developmental Disabilities, Memphis, USA

Allison L. Bruhn University of Iowa, Iowa City, USA

D. Tichelle Bruntmeyer University of Missouri, Columbia, USA

Matthew K. Burns Department of Educational Psychology, University of Missouri, Missouri, USA

University of Missouri, Minneapolis, USA

University of Missouri, Columbia, MO, USA

Judith J. Carta University of Kansas, Kansas City, KS, USA

Nathan H. Clemens Texas A&M University, College Station, USA

Robin S. Codding University of Massachusetts-Boston, Boston, USA

Melissa Coolong-Chaffin University of Wisconsin—Eau Claire, Eau Claire, USA

University of Wisconsin, Eau Claire, USA

W. Alan Coulter TIERS Group, Human Development Center, LSU Health Sciences Center—New Orleans, New Orleans, LA, USA

Stanley L. Deno Department of Educational Psychology, University of Minnesota, Minnesota, USA

University of Minnesota, Minneapolis, MN, USA

Christopher M. Dudek Rutgers University, New Brunswick, USA

Danielle N. Dupuis Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA

Lynn Edwards University of Minnesota, Minneapolis, USA

William P. Erchul Arizona State University, Tempe, AZ, USA

Jennie L. Farmer Tampa, USA

Amanda Fields Luverne Public Schools, Luverne, USA

Jack M. Fletcher Department of Psychology and Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, TX, USA

Barbara R. Foorman Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

Jeremy W. Ford Boise State University, Idaho, USA

Rachel Freeman University of Minneosota, Minneapolis, USA

Douglas Fuchs Peabody College of Vanderbilt University, Nashville, USA

Vanderbilt University, Nashville, TN, USA

Lynn S. Fuchs Vanderbilt University, Nashville, TN, USA

Kristin A. Gansle Louisiana State University, Baton Rouge, LA, USA

Brandy Gatlin Florida State University, Tallahassee, FL, USA

Maribeth Gettinger University of Wisconsin, Madison, WI, USA

Kim Gibbons St. Croix River Education District, Rush City, MN, USA

Howard Goldstein University of South Florida, Tampa, FL, USA

Charles R. Greenwood University of Kansas, Kansas City, KS, USA

Luana Greulich Andrews University, Berrien Springs, USA

Gabriela Guerrero University of Kansas, Kansas City, KS, USA

Renee Guy Luverne Public Schools, Luverne, USA

Aaron Haddock University of California, Santa Barbara, USA

Matthew Hall University of Minnesota, Minneapolis, USA

Lori Helman University of Minnesota, Minneapolis, USA

Kiersten Hensley Minnesota State University, Mankato, MN, USA

David Hoppey Tampa, USA

John L. Hosp University of Iowa, Iowa, USA

Sally Huddle Iowa Wesleyan College, Iowa, USA

Shane R. Jimerson University of California, Santa Barbara, CA, USA

Department of Counseling, Clinical, and School Psychology, University of California, Santa Barbara, USA

Asha K. Jitendra Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA

Ruth Kaminski Dynamic Measurement Group, Eugene, OR, USA

Abbey C. Karich University of Minnesota, Minneapolis, USA

Milena A. Keller-Margulis University of Houston, Houston, USA

Elizabeth Kelley University of MIssouri, Minneapolis, MN, USA

David A. Klingbeil University of Wisconsin-Milwaukee, Milwaukee, USA

Joseph F. Kovaleski Indiana University of Pennsylvania, Indiana, USA

Matthew Lau Minneapolis Public Schools, Minneapolis, USA

Timothy J. Lewis University of Missouri, Columbia, USA

David J. Lillenstein Derry Township (PA) School District, Hershey, USA

Maura Linas University of Kansas, Minneapolis, MN, USA

Daniel M. Maggin University of Illinois at Chicago, Chicago, USA

Kathrin E. Maki University of Minnesota, Minneapolis, USA

Amelia S. Malone Vanderbilt University, Nashville, TN, USA

Doug Marston Minneapolis Public Schools, Minneapolis, USA

Ryan Martin University of Massachusetts-Boston, Boston, USA

Jennifer J. McComas University of Minnesota, Minneapolis, USA

Scott R. McConnell University of Minnesota, Minneapolis, MN, USA

Kristen L. McMaster University of Minnesota, Minneapolis, USA

Jeremy Miciak Department of Psychology and Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, TX, USA

Dawn Miller University of Kansas, Lawrence, USA

Barbara S. Mitchell University of Missouri, Columbia, USA

Paul Muyskens TIES, Saint Paul, USA

Minneapolis Public Schools, Minneapolis, USA

George H. Noell Louisiana State University, Baton Rouge, LA, USA

Stephanie Al Otaiba Southern Methodist University, Dallas, TX, USA

Southern Methodist University, Dallas, USA

Miriam Ortiz Southern Methodist University, Dallas, USA

David Parker ServeMinnesota, Minneapolis, USA

Linda A. Reddy Rutgers University, New Brunswick, USA

Amy L. Reschly University of Georgia, Athens, USA

T. Chris Riley-Tillman University of Missouri, Columbia, MO, USA

Kay Rosheim University of Minnesota, Minneapolis, USA

Timothy J. Runge Indiana University of Pennsylvania, Indiana, USA

Timothy Scholten Ossining Union Free School District, New York, USA

Ann C. Schulte T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe, AZ, USA

Reza Shahroozi University of California, Santa Barbara, USA

Edward S. Shapiro Center for Promoting Research to Practice, Lehigh University, Bethlehem, USA

Elisa S. Shernoff Rutgers University, New Brunswick, USA

Mark R. Shinn National Louis University, Chicago, USA

Benjamin Silberglitt TIES, Saint Paul, USA

Diana Socie University of California, Riverside, CA, USA

Melinda S. Sota Illinois, USA

Rachel Stein University of California, Santa Barbara, USA

Karen C. Stoiber Univerisity of Wisconsin, Milwaukee, WI, USA

Karla K. Stuebing Department of Psychology and Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, TX, USA

George Sugai University of Connecticut, Storrs, USA

Gerald Tindal University of Oregon, Eugene, USA

Joseph Torgesen Florida Center for Reading Research, Florida State University, Tallahassee, USA

Janet S. Twyman University of Massachusetts Medical School/Shriver Center and The Center on Innovations in Learning, NY, UK

Amanda M. VanDerHeyden Education Research & Consulting, Inc., Fairhope, AL, USA

Mike Vanderwood University of California, Riverside, CA, USA

Alisha Wackerle-Hollman University of Minnesota, Minneapolis, MN, USA

Liliana Wagner Department of Educational Psychology, University of Georgia, Athens, Georgia

Jeanie Wanzek Florida Center for Reading Research, Florida State University, Tallahassee, USA

Jeanne Wanzek Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

Florida State University, Tallahassee, FL, USA

Caryn S. Ward University of North Carolina, Chapel Hill, USA

Jennifer Wilson University of Minnesota, Minneapolis, USA

Holly S. Windram Hope Network, Michigan Education Corps, Grand Rapids, USA

Myeongsun Yoon Texas A&M University, College Station, USA

From Response to Intervention to Multi-Tiered Systems of Support: Advances in the Science and Practice of Assessment and Intervention

Shane R. Jimerson, Matthew K. Burns and Amanda M. VanDerHeyden

The 2004 reauthorization of the federal *Individuals with Disabilities Education Act* (IDEA 2004) permitted a choice when diagnosing a specific learning disability (SLD) between (a) demonstrating continued poor performance when the student is provided with a research-based intervention and (b) conducting an assessment to demonstrate a pattern of strengths and weaknesses, which could include ability–achievement discrepancy (PL 108–446, Part B, Sec 614(b)(6)(b)). The federal regulations still required, as they *always* have, documentation of low achievement, a determination that the student's academic deficiencies were not the result of other disabilities or situational factors, and that learning deficiencies were not the result of insufficient instruction. The 2004 federal provision also allowed for response to intervention (RTI) to be used as part of the SLD identification process. Thus, an idea that had been discussed for years suddenly became a part of special education regulations (Gresham 2007).

Shortly after RTI was included in federal special education regulations, many school personnel realized that they did not have effective assessment tactics to evaluate these criteria when determining eligibility. The research base for RTI provided a set of codified tools and tactics that improved decision accuracy and student learning and made apparent when alternative causes of poor achievement could not be ruled out during the eligibility process (e.g., given instruction, the student's performance improves so poor instruction as a cause of low achievement cannot be ruled out). Better procedures and tactics to directly assess and rule out inadequate instruction as a cause of low achievement both enhanced decision accuracy and improved student learning. Today, many education professionals recognize that RTI involves universal screening, evidence-based instructional programming and curricula, routine progress monitoring of all students, increasingly intensive supplemental support and intervention for struggling students, and effective teaming practices; indeed, there have been many advances since 2004.

Policy documents and federal legislation governing educational service delivery over the past two decades have shifted to reduce the gap between instruction and evaluation. Recently, Kovaleski et al. (2013) discussed how to use RTI to determine eligibility for special education under the category of SLD and described the language used in Senate reports leading up to the passage of IDEA 2004 and No Child Left Behind (NCLB) 2004 as mutually referential. The revisions to IDEA were intended to advance the learning of all struggling students via high-quality, evidence-based instruction, to conduct effective primary prevention through regular screening and intervention support in general education, and to

S. R. Jimerson (✉)
University of California, Santa Barbara, CA, USA
e-mail: Jimerson@education.ucsb.edu

M. K. Burns
University of Missouri, Minneapolis, USA

A. M. VanDerHeyden
Education Research & Consulting, Inc., Fairhope, AL, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_1

permit more accurate identification of children requiring special education under the SLD label.

Contemporary educational policy language is aspirational and geared toward delivering results for all students and especially for students who might otherwise struggle academically and experience academic failure. Educational systems have shifted to value consequential validity (Kane 2013; Messick 1995) and to focus on diagnostic markers and processes that produce value for students who are made eligible for services. Scientifically informed and converging efforts in general and special education highlight the need and opportunity for enhanced prevention efforts directed at improving the proficiency of all students, frequent monitoring of and enhancement of the adequacy of the instructional system, emphasizing selection of evidence-based instructional strategies, fine-tuning of identification and eligibility decisions, and enhanced intensity of instructional offerings in special education. Identifying systemic and individual child needs, subsequently providing targeted interventions, and monitoring reduced risk and student growth (the core tenets of RTI) which were only emerging in 2006 when the first edition of this book was published, have now firmly arrived as best practices in enhancing student achievement.

Current Practices to Enhance Student Achievement

When the first edition of this book was written, RTI referred to a system of loosely grouped techniques accomplishing screening, progress monitoring, intervention delivery, and data-driven decision-making. RTI, with some variability, was being widely used and studied in several states. Currently, all 50 states encourage RTI for prevention purposes and a large and growing number of states allow the use of RTI for the identification of learning disabilities (Fuchs and Vaughn 2012; Zirkel and Thomas 2010) (for examples of state RTI initiatives, see Table 1).

Since the first edition of this book appeared, there has been a virtual explosion of RTI practices such that RTI structures and tenets are now considered foundational to best practices in school psychology (Ysseldyke et al. 2006) and consultation in schools (Reschly and Reschly 2014). RTI has been the focus of professional development efforts in nearly every state (Spectrum K–12 School Solutions 2010). The US Department of Education has funded resource and technical assistance centers (e.g., National Center on Response to Intervention, National Center on Intensive Intervention, National Center on Progress Monitoring, National Center on Early Childhood Response to Intervention) and has prepared practice guides on the use of RTI in schools via the What Works Clearinghouse at the Institute of Education Sciences (Gersten et al. 2008). Advocacy groups (National Center for Learning Disabilities) and professional educational associations (National Association of State Directors of Special Education, Chief State School Officers' Association, National Association of School Psychologists, Council for Exceptional Children) have given attention to RTI as a foundation in serving the educational needs of all students, including students who are vulnerable to failure. RTI processes typically include a data-based framework for decision-making, use of a problem-solving process across all levels of the system, and a team-based approach for leading, planning, and evaluating intervention effects (Hawken et al. 2008). Common core standards that are stable and rigorous have been adopted by most states and are consonant with the aspirations of NCLB and IDEA that all students master essential learning objectives by the conclusion of their formal education in the USA. Perhaps most notably, RTI implementations have merged with school improvement or school reform efforts and have been retitled multi-tiered systems of support (MTSS).

Several changes have occurred in K–12 schools as a result of research and policy changes. Curricula have emerged that incorporate assessments to evaluate skill mastery and learning progress and provide lesson supplement plans to assist teachers in layering more intensive instruction to improve student learning (O'Connor et al. 2005). Universal screening systems, once relatively rare and researcher-developed, are now

Table 1 Examples of state response to intervention and multi-tiered system of support initiatives

Alabama—https://www.alsde.edu/sec/isvcs/Pages/rti-all.aspx
Alaska—http://www.alaskacc.org/prodev/129
Arizona—http://www.azed.gov/special-education/state-initiatives/rti-mtss/
Arkansas—https://arksped.k12.ar.us/documents/special_show_08/Response_to_Intervention_Making_It_Work_in_Your_School.pdf
California—http://www.cde.ca.gov/ci/cr/ri/
Colorado—http://www.cde.state.co.us/mtss
Connecticut—http://ctserc.org/s/index.php?option=com_content&view=article&id=318%3Aresponse-to-intervention&catid=112%3Asrbi&Itemid=110&2fa6f942252db2ec6c621fe255459617=84461726d2f834623277c28b5fef3719
Delaware—http://www.doe.k12.de.us/infosuites/staff/profdev/rti_2014/
Florida—http://www.florida-rti.org/floridamtss/index.htm
Georgia—http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Curriculum-and-Instruction/Pages/Response-to-Intervention.aspx
Hawaii—http://www.hawaiipublicschools.org/Pages/home.aspx
Idaho—http://www.sde.idaho.gov/site/rti/
Illinois—http://www.isbe.state.il.us/rti_plan/default.htm
Indiana—http://www.doe.in.gov/student-services/student-assistance/best-practiceresearch-based-prevention-and-intervention
Iowa—https://www.educateiowa.gov/multi-tiered-system-supports-mtss
Kansas—http://www.ksde.org/
Kentucky—http://education.ky.gov/educational/int/ksi/Pages/default.aspx
Louisiana—http://www.louisianabelieves.com/
Maine—http://www.maine.gov/doe/rti/
Maryland—http://marylandlearninglinks.org/2502
Massachusetts—http://www.doe.mass.edu/sped/mtss.html
Michigan—http://miblsi.cenmi.org/Home.aspx
Minnesota—http://education.state.mn.us/MDE/EdExc/StandImplToolkit/Exploration/CriticalCompMTSS/index.html
Mississippi—https://www.mbaea.org/en/multitiered_systems_of_support_mtss/
Missouri—http://mimschools.org/
Montana—http://opi.mt.gov/Programs/SchoolPrograms/rti/index.html
Nebraska—http://rtinebraska.unl.edu/
Nevada—http://www.doe.nv.gov/SPED_Response_Intervention/
New Hampshire—http://www.education.nh.gov/innovations/rti/index.htm
New Jersey—http://www.state.nj.us/education/
New Mexico—http://www.ped.state.nm.us/rti/index.html
New York—http://www.nysrti.org/
North Carolina—http://www.learnnc.org/lp/pages/6880
North Dakota—http://www.dpi.state.nd.us/speced1/personnel/implement.shtm
Ohio—https://education.ohio.gov/Topics/Special-Education/Federal-and-State-Requirements/Procedures-and-Guidance/Evaluation/Instruction-and-Intervention-Supported-by-Scientif
Oklahoma—http://www.ok.gov/sde/oklahoma-tiered-intervention-system-support-otiss
Oregon—http://www.ode.state.or.us/search/page/?id=315
Pennsylvania—http://www.pattan.net/category/Educational%20Initiatives/Response%20to%20Instruction%20and%20Intervention%20%28RtII%29
Rhode Island—http://www.ride.ri.gov/InstructionAssessment/InstructionalResources/ResponsetoIntervention.aspx
South Carolina—http://ed.sc.gov/agency/programs-services/173/ResponsetoInterventionRTI.cfm
South Dakota—http://doe.sd.gov/oess/sped_RtI.aspx
Tennessee—http://www.tennessee.gov/education/instruction/rti2.shtml
Texas—http://www.tea.state.tx.us/index2.aspx?id=2147500224
Utah—http://www.updc.org/umtss/
Vermont—http://education.vermont.gov/educational-support-system

Table 1 (continued)

Virginia—http://www.doe.virginia.gov/instruction/virginia_tiered_system_supports/response_intervention/index.shtml
Washington—http://www.k12.wa.us/RTI/
West Virginia—http://wvde.state.wv.us/osp/RtIOSP.html
Wisconsin—http://rti.dpi.wi.gov/
Wyoming—http://edu.wyoming.gov/rti/

commonly used in many schools (Kettler et al. 2014).

Research and implementation efforts in RTI have also made apparent some of the threats to validity that certainly have existed for many years and remain pertinent today (e.g., intervention implementation integrity). The criterion that was new to the 2004 reauthorization and 2006 federal regulations was criterion 2 (use of RTI vs. use of cognitive battery of assessment to document strengths and weaknesses) and has been the focus of some controversy and debate (Colker 2013).

Context for the Current Edition

In the first edition of this book, contributing authors highlighted the historical emergence of RTI as a practice for serving struggling students, reducing student risk, increasing student achievement, and generating data that could permit earlier and more accurate identification of students in need of special education services. The emergence of RTI has been described as a logical evolution of best practices in education and in school psychology (Reschly 1988). Seminal scholarly articles, policy reports, and legislative language note several catalysts in the emergence of RTI as both an eligibility determination and student achievement improvement tool, many of which were highlighted in chapters in the first edition of this book (e.g., Gresham 2007; Kratochwill et al. 2007). First, a substantial research base scientifically discredited traditional SLD diagnostic practices as lacking validity and potentially causing greater harm than benefit to students as a result of misdiagnosis and weak outcomes for this group of vulnerable students. Second, concerns about general educational achievement in the USA, especially in the area of reading, raised concerns about the adequacy of instructional practices to establish minimal proficiencies for most students. Third, the rapid proliferation of the SLD diagnosis in concert with relatively weak evidence that the diagnosis led to specialized interventions that produced meaningful gains for the students receiving the diagnosis (Kavale and Forness 1999) raised questions about the value of the eligibility decision for students and whether eligibility and subsequent services caused these students to experience better outcomes than they would have experienced without eligibility. Fourth, the availability of short-term brief assessments to assess the overall health of the educational program, to monitor learning progress of students, and to permit midstream adjustments to instruction to improve learning were widely studied and readily available to educators. Fifth, more rigorous research evaluations and research meta-analyses and syntheses identified effective interventions that could be deployed to prevent failure for students who were experiencing academic risk.

In the 8 years since the first edition of this book, RTI implementation has grown and evaluation efforts show some promising findings. For the first time in the existence of the category, eligibility for special education under the category of SLD has declined. In 1980, approximately 10 % of students aged 3–21 years were diagnosed with an SLD, and that number increased each year until it reached 13.8 % in 2004–2005, but it then started to decrease for the first time since SLD was included in IDEA to 13.7 % in 2005–2006, 13.6 % in 2006–2007, 13.4 % in 2007–2008, to 13.0 % in 2010–2011 (US Department of Education, National Center for Educational Statistics 2013). In the first edition of the book, authors noted a need for more rigorous evaluations of interventions, study of RTI practices in preschools and secondary settings, use of RTI practices to promote learning in content areas other than reading. In the 8 years since the first edition of this book, rapid progress has been made in identification of effective interventions, development of practices and tools for use in preschool and secondary settings, and in mathematics.

In this second edition of the handbook, chapters describing the evolved and more rigorous research base for which a need was noted in 2007 are presented. First, chapters are organized according to foundations of science and practice, which provide chapters regarding the basic and applied research that led to RTI. Second, chapters are organized according to assessment, problem analysis, and intervention for each of the three tiers. The chapters that summarize practice as contemporary implementation science to bring in information from a relatively new branch of professional knowledge, but then also provided examples of novel applications and effective practices with chapters regarding contemporary issues and effective contemporary practices are also included. Each chapter was required to include some data to support their claims. Some of the chapters include experimental or quasi-experimental designs and some include surveys or case studies, but all of the conclusions are supported with data. Finally, in the authors' efforts to advance both science and practice, each chapter includes a table that highlights the implications for practice.

Future Direction and Challenges

Research outlined in this book and previously conducted highlights potential threats to validity of decisions made with RTI data, could negatively affect student outcomes, and present opportunities for future research. Over-assessment is a near-ubiquitous reality in most school systems. Many children are exposed to multiple early literacy screening measures that tend to have highly correlated scores and do not offer improved accuracy of decision-making as a result of their use. The collection of too much data complicates data interpretation and can introduce delays and error into the decision-making process. Collection of assessment data is resource intensive and comes at a direct cost to available instructional time. Additionally, very liberal cut scores designed to limit false-negative errors at screening create inflated numbers of students who appear to require intervention. Inefficient cut scores are burdensome to the system, complicate intervention support and deployment, and often compromise intervention effects. Research translating and supporting smarter screening procedures are needed to enhance efficiency of assessment and decision-making within RTI. Second, some of the criticisms levied against problem-solving teams (e.g., inconsistency in implementation, not following a problem-solving framework; Burns et al. 2005) are also applicable to grade-level teams, professional learning communities, and other school-based teams. In fact, although the research regarding problem-solving teams suggests inconsistent results, the effectiveness of Professional Learning Communities (PLCs) has yet to be thoroughly examined in the extant research. Finally, more research is needed to ensure better integrity of interventions, decision-making rules, and problem-solving processes.

Conclusions

Educational practices related to RTI continue to be modified, amidst an increasing array of resources that synthesize essential knowledge regarding the conceptual and empirical underpinnings of RTI and actual implementation. The IDEA legislation and many RTI initiatives during the past two decades continue to serve as a catalyst for further efforts and future scholarship to advance understanding of the science and practice of assessment and intervention at school. The *Handbook of Response to Intervention: The Science and Practice of Multi-Tiered Systems of Support* (Jimerson et al. 2016) provides a collection of chapters that address essential aspects and aim to enhance the future developments of RTI, toward facilitating enhanced student outcomes.

Whereas the roots of RTI are discernible in a research base that stretches back over the past 40 years in the areas of behavior analysis, precision teaching, direct instruction, curriculum-based assessment, measurement, and evaluation, and effective teaching, RTI remains an evolving science of decision-making. It is the authors' intent that this second edition of the *Handbook of Response to Intervention: The Science and Practice of Multi-Tiered Systems of Support (2016)* advance both science and practice, and enhance the lives of the children they serve.

References

Burns, M. K., Christ, T. J., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *NASP Communiqué, 34* (3), 1.

Colker, R. (2013). *Disabled education: A critical analysis of the Individuals with Disabilities Education Act.* New York: NYU Press.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45,* 195–203.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide*. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/. Accessed 15 Nov 2004.

Gresham, F. M. (2007). Evolution of the RTI concept: Empirical foundations and recent developments. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *The handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Hawken, L. S., Vincent, C. G., & Schumann, J. (2008). Response to intervention for social behavior: Challenges and opportunities. *Journal of Emotional and Behavioral Disorders, 16,* 213–225.

Individuals with Disabilities Education Improvement Act (2004). *Public Law 108–446 108th Congress*. Online at http://idea.ed.gov/download/statute.html. Accessed 15 Nov 2004.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.). (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.). New York: Springer Science.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review, 42,* 448–457.

Kavale, K. A., & Forness, S. R. (1999). *Efficacy of special education and related services*. Washington, DC: American Association on Mental Retardation.

Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (2014). An introduction to universal screening in educational settings. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 3–17). Washington, DC: American Psychological Association.

Kovaleski, J., VanDerHeyden, A. M., & Shapiro, E. S. (2013). *The RTI approach to evaluating learning disabilities*. New York: Guilford Publications.

Kratochwill, T. R., Clements, M. A., & Kalymon, K. M. (2007). Response to intervention: Conceptual and methodological issues in implementation. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 25–52). New York: Springer Science.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5–8.

O'Connor, R. E., Fulmer, D., Harty, K., & Bell, K. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and child outcomes. *Journal of Learning Disabilities, 38,* 440–455. doi:10.1177/00222194050380050701.

Reschly, D. J. (1988). Special education reform: School psychology revolution. *School Psychology Review, 17,* 459–475.

Reschly, A. L., & Reschly, D. J. (2014). School consultation and response to intervention: Convergence, divergence, and future directions for research and practice. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 495–512). New York: Routledge.

Spectrum K-12 School Solutions. (2010). *Response to intervention adoption survey 2010*. Eagan: Scantron.

U.S. Department of Education, National Center for Education Statistics. (2013). *Digest of education statistics, 2012* (NCES 2014–015). Washington, DC: author.

Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., Rosenfield, S., & Telzrow, C. (2006). *School psychology: A blueprint for training in practice III*. Bethesda: National Association of School Psychologists.

Zirkel, P. A., & Thomas, L. B. (2010). State laws and guidelines for implementing RTI. *Teaching Exceptional Children, 43*(1), 60–73.

Part I
Foundations of Science

Data-Based Decision-Making

Stanley L. Deno

The use of data to make decisions is both central and basic in the response to intervention (RTI) approach. While data are central to effective RTI, the procedures for basing RTI decisions on data are complicated and varied. The purpose of this chapter is to consider core ideas of data-based decision-making in RTI, to provide a perspective on issues related to data-based decision-making, and to recommend procedures for maximizing success in data-based decision-making.

Why Data-Based Decisions? A Personal Story

Early in the my tenure as a faculty member at the University of Minnesota, I was fortunate to receive funding to develop a collaborative training program for special education teachers in a nearby Minneapolis elementary school (Seward Elementary). A primary feature of that project involved spending my days in the school to help create a noncategorical special education resource program by disestablishing two segregated special classes. At the time, the setting was referred to as a university field station, and my role included working there each day with six to eight university students. The goal for the university students was to learn how to function in, what was then, a new special education program model in which students with high-incidence disabilities would spend most of their days in general education classrooms rather than special classes and receive supplementary instruction from resource teachers.

The role and procedures for functioning as special education resource teachers in such a setting were, as yet, undeveloped. And so the author's job was to develop both program and the role requirements for the resource teachers in training. The primary goal of our work there was to create a supportive academic program for all special education students that would enable them to acquire basic academic skills in reading, writing, and arithmetic. Since the students, now integrated, presented significant management challenges for their teacher, it was also necessary to develop the capacity of the teachers to address the social behavior of the special education students. In the approach to teacher education, each university student assumed responsibility for the programs of several of the Seward special education students who were now integrated into general education classrooms. My role was to provide direction and support for the university students as they assumed responsibility for designing the special education students' programs and that, typically, included supplementary tutoring.

Moving ahead in this project, the decision that was most challenging for me was the type of instructional program that the university students would use in attempting to increase the basic skills

S. L. Deno (✉)
University of Minnesota, Minneapolis,
MN 55455-0364, USA
e-mail: denox001@umn.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_2

of their special education students. I soon discovered that, while I had my ideas about what they should do, the students were taking classes from other university professors who were recommending approaches to teaching that were quite different from what I was recommending and different from one another. The challenge that I faced was how to resolve this dilemma—teachers in training who were getting competing perspectives on how best to teach struggling students.

Most of what is written in the remainder of this chapter emerged from that early experience in trying to provide professional training when competing ideas and uncertainty exists regarding how best to solve the problems confronted by struggling students. *The decision then, as now, was to be sure everyone became committed to the problems to be solved rather than the solutions they were going to use.* In effect, I wanted the goals the teachers in training would be using to be predetermined and specific and for those goals to drive their work. I wanted the means for attaining those goals to be free to vary so that the teacher trainees were motivated to integrate and use what they considered the best approach they could create from what they had learned from their various professors and from their experience. I also wanted them to be free to modify the programs that they created when those programs seemed not to be effective.

To create the context for this approach to intervening with struggling students, it was clear that two critical needs existed. First, to have a way to establish the determinacy of the goals; and, second, to have procedures for evaluating whether or not instructional programs were leading to the attainment of those goals. It was assumed that with these two components in place, to unleash the creative energy and motivation of the teachers in training without placing their special education charges in academic jeopardy would be possible.

Data-Based Program Modification

The approach that emerged from the early work at Seward Elementary School was called "data-based program modification" (DBPM) and was detailed in a book of the same name (Deno and Mirkin 1977). DBPM was conceived then, as now, as a sequence of decisions that are made when modifying a student's program and a parallel set of evaluation activities that are intended to increase the likelihood that program modification would be successful. The evaluation procedures were designed to inform the decisions to be made, and the central feature of the evaluation procedures used for each decision was data collection. In the elaborated form of DBPM, each decision made includes specific measurement, evaluation, communication/collaboration, and consultation activities intended to improve the outcome of the decision.

DBPM Assumptions and Current Considerations

Presented below are the five assumptions in the DBPM manual that provide the basis for the approach. They are as important now as they were then, but the original wording has been modified to clarify their relevance to RTI. Each of the assumptions is discussed as it relates to data-based decision-making in RTI.

Assumption #1 The effects of program changes/interventions for individual students are typically unknown and should be treated as hypotheses.

With the widespread emphasis on selecting evidence-based interventions in RTI, it is sometimes easy to forget that the evidence of effectiveness rests on outcomes for groups rather than individuals. Beyond that, in virtually all experimental comparisons, there is overlap between the distributions for treatment and comparison groups. In a practical sense, this means that some students receiving the more effective, evidence-based intervention actually had the same, or poorer, outcomes as students in the less effective intervention. What this means for implementing RTI interventions is that, depending on the effectiveness of the existing program, one can have confidence that the evidence-based intervention is likely to boost the progress, in general, for a group of students, but it cannot be said that will be the case for every student in the group. One needs to be mindful, then, that as an evidence-based intervention is implemented that one cannot predict for whom the intervention will work, and for whom it will not.

Assuming that the purpose of RTI is to improve the likely outcomes for all students, the fact that one cannot predict with certainty the students who will benefit from an intervention means that, no matter how strong the evidence for an intervention, careful attention must be paid to the effects for individual students. This is sometimes difficult to accept when one is working hard on assuring the fidelity of an intervention, but it is a reality that must be addressed. It also forces on us the prospect of having to create and manage a more differentiated program at each tier of our RTI model in an effort to be responsive to the progress of all students. Complete individualization is not logistically feasible, of course, but placing students in programs for most of the school year without being responsive to individual intervention effects would be unacceptable.

Assumption #2 Single-case research designs with repeated measurement data are well suited for testing intervention hypotheses.

Assuring that one is using data to monitor intervention effects for individual students requires practical program evaluation procedures that provide the best possible evidence for determining whether a program is meeting its goals for individual students. At this point in the development of RTI models, as has been the case in the past, the best available evaluation procedures are those that involve repeated measurement of individual student performance. Repeated measurement of performance across time produces a record (a time series) depicting the extent to which a student is progressing satisfactorily toward performance goals (Skiba and Deno 1984). These data can be configured and displayed graphically to provide an easily interpretable picture of a student's past, present, and likely future performance (see Fig. 1, for example). Even though repeated measurement of performance produces a data record that enables evaluation of intervention effects, it is important to recognize that many questions exist regarding the best approach to implementing a repeated measurement system. One seemingly simple question is how frequently student performance must be measured to provide a sufficiently reliable and valid database for

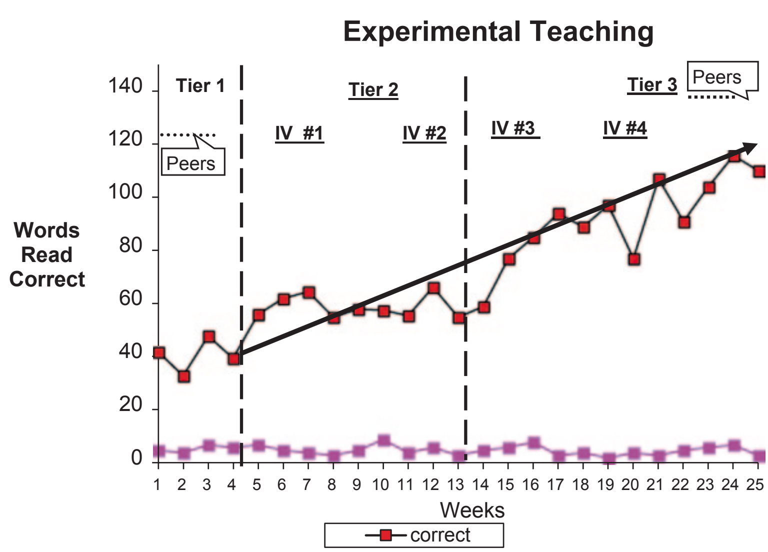


Fig. 1 Experimental teaching

testing individual intervention effects (Jenkins et al. 2005; Jenkins et al. 2009). This includes both practical and technical considerations. Since the system is to be logistically feasible, keeping the number of measurements to the minimum necessary to provide reliable and valid data is practically important. At the same time, making a judgment about whether a student is making sufficient progress as a result of an intervention requires a conclusion based on a student's rate of growth before and after the intervention has been implemented. Unfortunately, research on estimating growth from repeated measurement data has not sufficiently answered the question of how many measurements are required to reliably estimate growth (Ardoin and Christ 2009).

Assumption #3 Modifications in the general education program for individual students are hypotheses whose effects should be empirically tested.

It goes without saying that the best possible RTI model rests, first, on a core educational program for all students (tier 1) that is likely to be maximally effective. It should be the goal of every RTI model to minimize the number of students who will be identified as needing a tier 2 intervention. Having said that, even the most effective tier 1 programs will contain students whose needs are not being sufficiently addressed in the core program and who are at risk for academic difficulty. In such cases, students will be placed in an alternative or supplementary program (tier 2) focused on reducing their risk of failure. As this is done, it is important to be mindful that any program differing from the general education program rests, in fact, on the hypothesis that the alternative program will benefit students placed in that program. As stated earlier, no program is going to benefit all of the students all of the time. Thus, placement in an alternative program requires careful monitoring of program effects.

It is important to be reminded of this need to monitor the effectiveness of tier 2 programs because progress monitoring requires a commitment of time and resources that is tempting to avoid. In some cases, it might seem that a student should experience the alternative program for half or, perhaps, an entire school year to determine its effects. In such a case, it might seem reasonable to wait until the end of the program to evaluate its effects. Waiting for such a long duration to test the intervention effects, however, risks the very real possibility that some students are not benefitting from the intervention and that a large segment of the school year might pass before changes are made to those students' programs. A responsive RTI model includes procedures for frequently evaluating intervention effects throughout the school year rather than waiting for most of a school year to evaluate intervention effects.

Assumption #4 To use single-case designs for testing intervention hypotheses requires specification of the "vital signs" of educational development that can be routinely obtained in and out of school.

Forty years ago, simple and direct measures of student performance that could be routinely obtained to monitor the academic health of young students did not exist. Student performance was informally appraised by teachers through daily work with their students and occasional formal assessment was accomplished through schoolwide standardized achievement testing. Accountability requirements were virtually nonexistent and no requirement existed to assure that students given remedial assistance actually benefitted from that assistance.

In the intervening years, substantial amounts of effort and resources have been devoted to developing technically adequate and logistically feasible measures of student academic growth. Now, a menu of options exists from which educators can select progress-monitoring measures that enable "vital sign" monitoring and can serve as the basic data for making RTI programs responsive to student performance (National Center on Response to Intervention 2012). A review of the National Center on Response to Intervention "tools chart" reveals that different approaches have been developed to monitor student progress. The center's evaluation of those measures is intended to establish that all included measures can, potentially, be used within an RTI model. As is also clear from the chart, however, most of the progress-monitoring measures are based on the general outcome measurement (GOM) approach (Fuchs and Deno 1991) on which the curriculum-based measurement (CBM) approach is based (Deno 1985).

Research on this approach spanning more than 30 years has supported the technical adequacy and the utility of the GOM/CBM approach for generating the student performance data necessary for making a wide range of programming decisions (Espin et al. 2012). A very large number of juried empirical and nonempirical articles on GOM/CBM have been published in the professional literature of school psychology and special education (Speece 2012), and many states include those measures in their statewide testing. Nevertheless, several important issues on the use of these procedures require consideration.

The first issue to consider is whether the technical adequacy of GOM/CBM data is sufficient for making the high-stakes decisions that are required in RTI. CBM was developed as a tool for teachers to use in formatively evaluating their instruction with students who were academically disabled. The original research hypothesis to be tested using CBM data was that teachers could increase student achievement by collecting CBM data and using it to make decisions about whether to continue or change their instruction. Initial research supported this hypothesis (Fuchs et al. 1984). Since that early work, GOM/CBM has been increasingly used to make a wide range of decisions, including screening for alternative programs, writing annual individualized education program (IEP) goals, eligibility for special education services, evaluating program effects, and as evidence in legal proceedings (Espin et al. 2012).

As has been pointed out by others (Davis et al. 2007), placement of students at different levels in an RTI model is a high-stakes decision as it is practiced in most schools. CBM/GOM procedures are often used in the screening of students to identify those at risk for failure. While GOM/CBM procedures have been, and can be, adapted to a variety of uses, the extension to making decisions where the stakes are higher than that for which the measures were initially developed must be done carefully to assure that the data are valid for the decisions being made. Evidence exists that when using only GOM/CBM data to screen grade 1 students to identify those who will be placed in tier 2, a significant number of errors are made (Jenkins et al. 2007). A better approach is to use GOM/CBM as one measure in a battery of assessments (Compton et al. 2006; Jenkins and O' Conner 2002). Doing so minimizes classification errors to a degree that both assures that students who require intervention will receive it and that resources will not be spent on students who will succeed without intervention.

Other issues related to using GOM/CBM in an RTI framework exist, as well. Repeated measurement of individual student academic performance creates the opportunity to measure student growth and current level over the relatively short durations of an intervention. When teachers are using the measures informally to aid in their judgments about how students are progressing, the stakes are relatively low. However, when growth is being used to determine whether students are benefitting from their current tier placement or whether a move to a more intensive level is required, this is a high-stakes decision. As with using GOM/CBM as a measure for screening, using the repeated measurement data it produces to estimate growth is fraught with technical problems that have not yet been adequately solved. Growth trends estimated from GOM/CBM tend to be quite inconsistent and unreliable, with large confidence intervals that make data interpretation difficult (Ardoin and Christ 2009; Christ and Coolong-Chaffin 2007). Despite the problems associated with making high-stakes decisions using GOM/CBM growth data, the GOM/CBM approach is one of the only viable alternatives for examining individual student growth since other achievement measures tend to be insensitive to change over relatively short durations (Marston et al. 1986; Skiba et al. 1986). Continuing research and development designed to improve the technical characteristics of progress measures is underway and promises to address this important issue (Christ and Ardoin 2009).

Assumption #5 Testing intervention hypotheses requires well-trained professionals capable of drawing valid conclusions about program effects from the "vital signs" data.

The fifth assumption focuses on the heart of data-based decision-making: the decision-maker's use of data to make intervention decisions. As was true in the mid-1970s, the key element in successful, data-based programs is the people responsible for organizing and implementing programs and using data to increase their effectiveness. In the early development of data-based programs using vital sign progress monitoring, the decision-makers were special education teachers using the data-based approach to intervene with individual special education students. As the data-based approach has become more widespread, the training and competence of an increased number of professionals has come into play. Now, classroom teachers, school psychologists, and program administrators, among others, are required to examine student performance data to make decisions ranging from individual student progress to program evaluation. Regrettably, collecting and interpreting student performance data has never been a major component of teacher education programs (Stiggins 1993). Almost certainly, the same might be said with regard to the training of school administrators. The one group of educational professionals who are highly trained in assessment is the school psychologist. Not surprisingly, school psychologists often become key personnel in data collection, management, and interpretation in RTI. Typically, however, their work is done at the program level, rather than the individual student level, simply because the ratio of school psychologist to students is limited. Consequently, teachers still remain at the front line in the interpretation of individual student data.

Since teachers bear the primary responsibility for monitoring and interpreting individual student progress data, it is important to consider how well they are able to do this task. Espin and colleagues recently examined the accuracy and quality of teachers' interpretations of the typical progress-monitoring graphs used in RTI (Espin et al. 2012). The results of their research provide both good and bad news. Teachers' interpretations were compared to the interpretation of experts, and while some of the teacher's interpretations were consistent with the experts, others differed widely. Espin and colleagues described the poorer interpretations as lacking coherence, specificity, and accuracy. Obviously, there is room for improvement in the interpretation of student progress graphs. Even as this point is made, however, one needs to remember that, at best, the inconsistency of most progress measures for modeling individual student growth means that even experts in interpreting such data are using a database that is insufficient for making high-stakes decisions.

Using Decision Rules to Make Intervention Decisions To take full advantage of the repeated measurement data derived from progress monitoring requires not only that the data be collected and summarized but also that those managing interventions be responsive to the data in an effort to build a more effective intervention. Unfortunately, even when progress-monitoring data are systematically collected throughout the course of the academic year, it is common to find that teachers might make only one or two deliberate changes to increase intervention effectiveness (Deno et al. 1988). The problem one confronts is how to assure that the data collected during intervention are used as much to evaluate and improve the intervention as they are to evaluate the student.

A Caveat: The Data Do Not Speak for Themselves An oft-heard fallacy is that "the data speak for themselves." This, of course, is not true. The data are only numbers without meaning until one interprets them. In a short article on this topic, Behn (2009) makes the point by using the familiar example of a glass of water filled to its midlevel. The optimist describes the glass as half full, while the pessimist claims it is half empty. Both are using the same data point, but each interprets it differently. When we collect data on student performance, the situation is precisely the same. Student scores are meaningless until we apply our perspective to those scores. Even as we summarize data using simple descriptive statistics like averages and percentages, we need to be aware that our history with those statistics tends to operate in such a way that we are immediately attributing meaning and significance to those numbers. When we screen students, the numbers that are generated become valuable to us when we apply a decision framework to those numbers for purposes of identifying those students who are progressing well and those for whom additional attention will be necessary. We need to be fully aware that the decision framework that is used rests on both assumptions and values regarding the importance of students performing at different levels. While it is useful to have available standards and cut scores to enable us to make decisions, we cannot be misled into thinking that the data speak for us. Carrying out system-level directives does not change the fact that we are accepting the values and the decision frameworks designed by others and, in so doing, we complicit in making the decisions.

Professional Expertise in Decision-Making As we consider the expertise required to make data-based decisions in RTI, a question to consider is, how to increase the likelihood that the decisions made improve the outcomes? A pervasive assumption is that the best possible approach to data interpretation and decision-making is one rooted in the collective judgment of experienced decision-makers. However, this assumption has been both questioned and contradicted in the clinical psychology literature. In his classic book, *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence,* Paul Meehl examined the accuracy of clinicians in predicting individual outcomes. As is true for educators, most judgments about the clients of clinical psychologists involve predicting future outcomes from a variety of data sources (records, interviews, test scores, etc.). Once the data from these sources is quantified, two options exist. The first option is for professionals and others with a vested interest in the outcomes to examine the data and make a prediction regarding the best treatment based on their interpretation of the data (i.e., clinical prediction). The other approach that Meehl examined was to use a formula created to make the prediction (i.e., a statistical prediction). In the latter case, once the formula has been created, that prediction decision is the product of the formula rather than the professionals interpreting the data. It is important to note that, even though a formula is being used, professional judgments can be included in the formula. However, in the statistical approach, the professional judgment does not determine the decision, the formula does. After reviewing the research where the two approaches were compared empirically, Meehl found that the evidence supported a statistical, rather than clinical approach, in making predictions. As might be expected, Meehl's book has become the center of controversy in the field of psychotherapy, but his conclusions remain supported after more than 50 years (Grove and Lloyd 2006).

Meehl's findings are important for developing data-based RTI programs that involve predicting student outcomes. The very nature of RTI involves a host of significant decisions, including identifying students who are at risk of continuing academic difficulty and who will benefit from intervention, the types of interventions that will best serve students, whether students are benefitting from intervention, whether students should be moved to a more intensified level of intervention, and so on. As one seeks to make data-based decisions, it would be well to consider whether there is an opportunity to use more statistical, rather than a clinical, approaches to making those decisions. Certainly, from the standpoint of accuracy, an argument can be made that we should try to do whatever we can to increase the accuracy of our predictions. From the standpoint of efficiency, using formula-based approaches could well reduce the amount of personnel time involved in making the many RTI decisions, potentially, increase responsiveness to student progress data. While using a formula-based approach might seem mechanistic and impersonal, as is discussed in the next section, such an approach is already being used by many practitioners using progress-monitoring data.

Decision Rules The use of formula-based approaches with progress-monitoring data is described in the literature as "data decision rules" (Ardoin et al. 2013). Teachers have been trained in the use of data decision rules for more than 30 years (White and Haring 1980), and decision rules were part of the effective intervention reported previously (Fuchs et al. 1984). Those decision rules are generally based on data in the progress-monitoring graph where current student performance is plotted against the desired line of progress based on students' year-end goals. The slope of that line of progress is established by drawing a line that connects students' initial level of performance to the level established in the goal (see Fig. 1). Basically, two different types of decision rules based on that progress line have been used to prompt teachers to change programs or goals (Ardoin et al. 2013). The first simply specifies that the intervention should be changed if a prespecified number of consecutive data points fall below the line. The second requires that the slope of student scores be estimated and that slope then compared to the slope of the desired line of progress. If the slope is flatter than the slope of the progress line, then the rule directs a change in the intervention. If the slope is similar to the slope of the line of progress, then the rule is to continue the existing intervention. If the slope of student scores exceeds that of the desired progress line slope, then the rule directs either raising the goal (thereby increasing the slope of the progress line) or moving to a less intense intervention.

As has already been discussed, trend estimates from progress-monitoring data tend to be very unreliable. To overcome this difficulty, simple paper-and-pencil procedures have been developed that improve agreement in trend estimates based on the existing data (Skiba et al. 1989; White and Haring 1980). Improving agreement across those interpreting the progress-monitoring data is important, of course, but if the data themselves are unreliable, agreement in interpretation will not increase the reliability of the database for the decision to be made. Indeed, as Ardoin and colleagues concluded in their review, the present state of the empirical evidence regarding the effectiveness of formula-based decision rules is insufficient to recommend any particular decision rule for guiding when to change or maintain an intervention (Ardoin et al. 2013).

As computers have become readily available to collect, store, and analyze student progress data, additional possibilities for using formula-based decision-making have become possible (Fuchs et al. 1989a). Now, instead of using paper-and-pencil procedures to estimate growth, computers can quickly provide consistent information about student performance. The fact that computer technology can be used to enhance teacher planning and improve student outcomes was established by Fuchs and colleagues more than 20 years ago (Fuchs et al. 1989a, b). Surprisingly, little has been done in the intervening years to build decision rules into computer-based systems. As a recent summary of the use of technology with GOM/CBM makes clear, web-based systems are now available to facilitate progress monitoring (Lembke et al. 2012). At the same time, the development and use of technology has focused more on data summaries and presentations and tend not to include decision rules designed to direct change in student programs. Perhaps this is a good thing, given the findings of Ardoin and colleagues that the evidence base for decision rules is limited. At the same time, the availability of computer technology should enable researchers and developers to design effective decision rules that can be incorporated into those now widely used web-based systems.

An alternative approach to using computer analyses of student performance to enhance intervention is illustrated in the work of Fuchs and colleagues (Fuchs et al. 2005). The focus of that work has been on adding diagnostic information to the progress monitoring so that teachers derive specific information about the profile of skills that students have mastered and those still needing attention. These data are then used to both group students with similar skill sets and to design interventions related to those skills. That group of researchers has been able to demonstrate that developing supplementary diagnostic data can indeed enhance instructional effectiveness. At the same time, collecting data on skills thought to contribute to developing generalized competence in reading, writing, and mathematics can also result in a shift in attention from dependent variables to independent variables.

Distinguishing Between Dependent and Independent Variable Data

One of the most basic ideas taught to beginning researchers is the distinction between independent and dependent variables. Independent variables are those factors that the researcher manipulates directly in an effort to determine whether those manipulations produce a change in a measured outcome. In educational interventions, the independent variables are those changes (interventions) made in students' programs to produce higher levels of achievement. In contrast, dependent variables are not directly manipulated by the researcher. They are the outcomes, or the effects, that enable determination of the power of the independent variables. This research, of course, is an exploration to determine whether cause-and-effect relationships exist between the independent and dependent variables. In education, the primary dependent variables are the achievement scores collected annually. Most often these are state standards tests or standardized achievement tests. In RTI, the dependent variables need to be more routinely available than the annual state achievement or standards tests so that judgments can be made more frequently regarding the effectiveness of the interventions. That is one of the primary reasons why progress-monitoring data are collected and, in RTI, the dependent variables are typically progress measures.

It takes only a short step to see that in our work we are continually involved in manipulating variables in the environment in an effort to positively impact student outcomes. Thus, teaching, in general, can easily be viewed in terms of independent and dependent variables; and RTI, in particular, can be viewed as applied research in which combinations of independent variables are sought to produce desired changes in the dependent, progress-monitoring data.

Why is this distinction between independent and dependent variables important for us when considering data-based decision-making, and how does it relate to our work? The answer is that high-quality RTI programs rest on collecting and using the best possible dependent variable data. All of the most important decisions made in implementing RTI are grounded in the assumption that the data used to make program change decisions are reliable, valid, and possess consequential validity (Messick 1989). Before we begin implementing our interventions, our independent variables, we must be sure that we have measures of student outcomes in place that provide us with the information necessary to evaluate intervention effects. Further, to the degree possible, we should strive to be sure that continuity exists in the dependent variable data used to make the decisions about the effectiveness of our programs. Consequently, to the extent possible, we should collect and use the same data to identify students who are academically at risk, to monitor their progress during intervention, to evaluate the effectiveness of the interventions that have been created, and to move students to a different level within the RTI framework. If students are identified as at risk by using a battery of measures, then it will be most useful if the ongoing dependent variable data are included in that battery and collected throughout intervention and program evaluation. Because all measures possess some degree of unreliability, using different measures for different decisions increases the likelihood that measurement error will be compounded and that inaccurate decisions will be made as a function of this compounding of error.

While dependent variable data can aid us in deciding whether a program is succeeding, it is not designed to provide information about potentially effective independent variables that might be included in an intervention. A simple analogy based on the vital signs is the thermometer. It is used regularly to measure body temperature, but it provides no information on what to do if the temperature is too high. Similarly, a bathroom scale can help us to monitor weight gain, but it does not provide direction regarding what to do should we seek to change our weight. When GOM/CBM is used as the primary dependent measure, like the bathroom scale, it does not direct us to techniques or interventions that might alter the GOM/CBM scores. That information must come from other sources.

In an effort to provide decision-makers with data that can be used aid in designing interventions, some programs incorporate criterion-referenced tests that are used in initial identification of students at risk. For example, a criterion-referenced math test might be used to determine skills students have not yet mastered so that practice on the skills might be included in an intervention for those students. Another example is an early reading assessment that identifies decoding skills that students have, and have not, mastered. It is important to emphasize, however, that even though data have been collected revealing lack of skill mastery, the impact of including practice on those skills remains a hypothesis that must be tested by determining whether teaching those skills results in changes in the dependent measures. In other words, do the students' scores increase as they learn the skills?

When data are collected on skills thought to be necessary for a student to become generally competent in reading, writing, or arithmetic, it is relatively easy to become primarily focused on the skills that have been identified and to lose track of performance on the dependent measures. When that occurs, it is easy to mistakenly base evaluations of program success on the extent to which the identified skills have been mastered rather than whether real change has occurred in the generalized competence sought. This happens most often when students are assessed with a measure that includes a variety of subtests that have associated benchmark scores.

Suppose, for example, an initial reading assessment is conducted that includes measures of phonemic awareness, letter naming, onset phonemes, blending and nonword identification, and that each subtest has a "mastery" benchmark score. Given this set of initial reading skills, one might create an intervention targeting the skills in the assessment. And, as one implements the intervention, it might seem sensible to evaluate intervention effects by collecting data on whether students are mastering the various skills targeted in the assessment. The question that must be answered in this case is whether skill mastery constitutes independent variable manipulation or whether they are the primarily dependent variables. The answer is that the data collected on mastery of those skills enable us to determine whether the instructional approach has been implemented, not whether the students have improved in their general reading competence.

To test the hypothesis that the students meeting benchmarks on specified pre-reading skills will become better readers requires not only that we measure their attainment of those skills but also that we use a measure that possesses validity as a general measure of reading. As the students achieve benchmark performance on the skills, we should be able to observe a corresponding increase in our dependent measure. Referring back to my personal experience in training preservice teachers, our first step in building a data-based program was to create the dependent variables, the measures of student performance that the trainees would use as a basis for testing the effects of their ideas about how best to teach their students. To apply research techniques in creating effective interventions, they needed to be free to use alternative independent variables when their initial efforts were not effective. Focusing on the dependent variable enabled them to think flexibly about alternatives that might effect change in the dependent variable, rather than being committed to particular intervention approaches.

It would not be surprising if many were to think that the reading skills identified in the above example would be the appropriate dependent measures in an early reading intervention. They might argue that these should be the outcomes toward which interventions are directed. To take that position, however, requires research that measures of these skills possess criterion validity as reading measures. A good example is the case that can be made for two other measures that might serve as the dependent measures of reading in first grade. Both isolated word recognition and reading aloud from text can be measured throughout the entire first-grade year and more closely represent actual reading. For that reason they, for example, would function well as dependent measures for evaluating intervention effects in beginning reading (Fuchs et al. 2004).

The point here is not to argue for or against any particular approach to teaching and measuring beginning reading, but to highlight the importance of carefully considering the kinds of data that will be used to make critical intervention decisions. And, in doing so, to be clear about whether the data one is collecting are part of the intervention approach (independent variable data) or are data that enable us to determine whether the intervention is actually effecting the outcome (i.e., the dependent variable) for which the intervention has been created. In a well-designed RTI model, dependent variables are the primary determiners of whether or not a student is at risk for reading difficulty and whether, through intervention, the student is improving as a result. Additional data might well be collected during screening, but these data are used primarily to increase screening accuracy and to inform decisions about the type of intervention that will be used.

Making Decisions for Individual Students

Screening If we are to use data to help improve the decisions, we should be sure to rely on probabilities derived from evidence-based research. That is, we should make judgments about performance based on knows has been learned from prior research with students in similar circumstances. In the case of using recommended benchmark scores from progress-monitoring data to establish year-end goals, those benchmark scores will most certainly be derived from prior research showing that students who achieve those scores are on track to succeed at passing a criterion (perhaps a state standards test). Implicit in those benchmark scores is the inference that it is appropriate to use the same score for other students who share the characteristics of those students from whose performance the benchmark score was derived. At one level, this is a very good approach to take since it provides guidance to us for making decisions about the likelihood of student success. Even as we use data in this way, however, we need to be mindful that the evidence-based predictions we make are based on probabilities derived from analyses of group data. In this example, the benchmark score is, in effect, a probability statement that says students, in general, who achieved the benchmark in the past were *likely* (but not certain) to succeed on the criterion. Further, this probability statement refers to what was true for the group of students whose data were analyzed. It is not a statement about any individual in the group. Even in a case where 90 % of the students who achieved the benchmark passed the criterion test—and 10 % of the students did not, we are unable to predict which students will be in the 90 % that pass and which will be in the 10 % that fail.

The point is we must realize that using data to select an intervention for an individual student actually involves making a prediction about what might be effective for that students based on probabilities derived from research with groups of students. We need to be mindful that the probabilities we use in making those predictions are derived from the research and apply to groups rather than to individuals. This issue becomes particularly salient when using cut scores to initially screen students for tier placement in RTI programs. In such cases, we need to maximize the probability of a successful prediction for individual students. Research on this topic has demonstrated that cut scores can be raised to increase the likelihood that students who exceed the score will pass the standards test, but doing so will result in an increase in the proportion of students who will be incorrectly identified as at risk (Jenkins et al. 2007). These are students who scored below the cut score, but did, in fact, pass the criterion test. The same thing is true at the low end of the distribution. It is possible to assure that students who fall below the cut score will, in fact, not pass the criterion test by setting a low cut score that allows predicting with assurance that everyone with a score below that low score fails the criterion test. The lower that cut score is set, however, the greater the likelihood that we will miss identifying some students because their screening score is higher than the cut score and they, in fact, fail the criterion test. This issue is referred to as the *sensitivity and specificity* of the classification scores. Sensitivity refers to whether the classification procedures we use identify *all* students who are at risk and specificity refers to whether those procedures identify *only* those who are at risk. We need to be aware that even the best data-based classification procedures will not be 100 % accurate; yet accurate identification is an important problem in using data to make placement decisions as it bears on the efficient allocation of resources. Our use of data should, as much as possible, assure that we identify only those who require intervention and that we do not use resources to intervene with students who will succeed without intervention.

Interventions Selecting evidence-based interventions provides us with another good example of the difficulty arising from using group data to make individual predictions. Interventions that have been identified as evidence-based in terms of their effectiveness typically have been carefully researched through experimental field tests. Experiments of this sort compare groups of students receiving an intervention with groups that do not. Judgment of intervention effectiveness is based on the degree to which one group, in general, performs better than the other. Once interventions have been completed, statistical comparisons are made of the average performance of the groups and the degree to which the scores of the two groups overlap. The results enable conclusions about whether an intervention has been more effective *for the group*. If so, the evidence base has been established for that intervention.

Even when experimental evidence of this sort is available to help make decisions about the interventions to use, one must recognize that the *evidence provides data on the effectiveness of an intervention for the group, but not for all individuals in the group*. When using research evidence to make predictions about individual students, we need to be aware that data bearing on findings for a group do not establish that the findings apply to every individual. We can never predict with certainty that a particular intervention will be effective for an individual student. This reality, of course, provides us with the need to carefully following up the effects of interventions with individuals through regular progress monitoring.

Data-Based Problem-Solving in RTI

A useful perspective on the purpose of RTI is that its goal is to solve problems that emerge from our efforts to enhance academic and social development through education. Interventions are created when the standard educational program being provided for all students is not working for some students. As simple and obvious as this seems, the idea that a problem exists that requires solution rests on several very important assumptions that must be explicitly clarified if effective data-based decisions are going to be made in RTI models. A problem-solving model useful in RTI has been considered elsewhere (Deno and Mirkin 1977, 1985, 1995, 2005) and will not be described in detail here. However, several key elements of that model are worth considering as they lay the basis for a data-based approach to RTI decision making. Those elements are (1) identifying a problem, (2) defining the problem quantitatively, and (3) evaluating intervention effects.

Identifying the Existence of a Problem Since the need for intervention rests on the assumption that a problem exists in the academic or social development of a student, consideration must be given to how the problem has been, or will be, identified. Until quite recently, the suggestion that students had problems was based largely on their teachers' judgments and expressions of concern. Parents and school administrators relied on teachers to raise concerns about the progress of students and, for the most part, accepted the judgments of those teachers. The assumption was, of course, that no one knew better how the students were progressing than their teachers, and that teachers would be in the best position to express concerns about students who seemed to be struggling. In general, when teachers identified students who they saw as having problems, parents were informed and plans were developed to try to help students overcome those problems. In such cases, problem identification could be viewed as a result of teachers holding expectations for performance that were not met by the students. A simple way to think about the nature of such problems is as discrepancies between *teacher expectations* and *student performance*.

With increased emphasis on the use of tests to establish teacher accountability, the emphasis in responsibility for identifying academic problems has shifted from individual teachers to a structured, system-based approach where performance standards are set by states and by school districts. In this context, students are identified as having problems because they are not meeting the test-based standards set by the state and the districts rather than the implicit standards held by their teachers. In this scenario, *problems are defined as discrepancies between state and district expectations and student test performance*.

The distinction between teacher-identified discrepancies and system-identified discrepancies might seem unimportant, but many and varied implications follow from the source that identifies performance discrepancies and why those discrepancies are considered important. When teachers identify student problems, the basis for that identification is, by its very nature, individualistic and often idiosyncratic. Teachers' interactions with their students are personal, and their views of who has a problem are influenced by those interactions. Further, their performance expectations are influenced by their personal biases related to student demographic characteristics such as gender and race and the performance, in general, of students in the schools where they teach. When teachers are in schools where student performance is generally low, it is natural for them to view the low performance of any one of their students as being consistent with the rest of the students in the school. In addition, when students are compliant and not creating management problems for the teachers, those teachers are less likely to see them as "having problems." In contrast, when districts set test standards, the role of the teacher in determining who the underperforming students are is eliminated along with the potentially biasing effects of teacher's views. Moving to test-based problem identification carries its own risks, however. When tests become the sole determiner of performance, the reliability and validity of those tests and the appropriateness of the standards set with those tests must be very high to minimize errors in identification of students having problems. As discussed previously, this issue is especially salient during the screening phase of RTI models where criteria must be established in such a way as to maximize the probability that students who are, in fact, at risk of academic failure are identified for intervention, and to minimize the probability that students who do not require intervention are not identified as at risk academically.

Another issue that arises when shifting problem identification from the teacher to the state standards tests is that those tests are typically summative and given toward the end of the school year. The result is that, like closing the barn door after the horse has already escaped, it is now too late to attempt to solve the problem during that academic year. Data-based efforts at problem-solving through RTI are intended to prevent or to reduce potentially serious problems and their negative consequences. As discussed previously, this means that data-based decision-making in an RTI framework requires identification of problems early in the school year and that data in addition to the annual state standards tests will have to be generated at the beginning of the school year to both identify and define the problems. The need for measures that can be used throughout the school year has led to a considerable amount of research on the predictive validity of both screening and progress-monitoring measures (Wayman et al. 2007; Jenkins et al. 2007; Speece 2012). While the results of this research have been encouraging, predictions over the course of an academic year are far from perfect. In addition, the technical characteristics of the screening and progress measures make accurate predictions difficult (Christ and Coolong-Chaffin 2007). This increases the difficulty of establishing effective procedures for identifying those problems to be addressed through RTI. Given the uncertain level of validity of current assessment procedures for making long-range predictions, one inescapable conclusion is that schools should continually conduct research on their problem identification procedures to assure their validity.

Defining Problems Quantitatively It is one thing to identify the existence of a problem, it is quite another thing to attempt to describe or define that problem with some precision. It is the case that test-based problem identification tends to obscure the fact that values, in the form of expectations for performance, are establishing the basis for viewing student performance as discrepant. Developers of RTI models need to recognize and make explicit the existence of those values. In developing an RTI model, we need to ensure that the tests being used to generate the data for problem identification are technically adequate. Too often, in the rush to establish accountability systems, states have created tests of unknown reliability and validity. Setting standards using such tests will serve only to both over- and under-identify those students who appear to be at risk for academic failure (Brandt 2012). If RTI is ultimately judged to be a successful approach to reducing school failures and improving student achievement, it will be because it has helped to increase the proportion of students who attain the standards derived from our cultural imperatives (Reynolds and Birch 1982). When those standards are codified as scores on tests, every effort must be made to assure that adequate evidence exists for using those tests and for the performance standards established on those tests. Quantitative definitions of the problems RTI is designed to solve begin with the expectations set by those standards.

When defining problems quantitatively, it is most useful to define the problems that have been identified using the same data that will be used to determine whether the intervention is working (formative evaluation) and, ultimately, has been effective (summative evaluation). Regrettably, this is not always possible since, with test-based based discrepancies, the cultural expectations most often are embedded in the scores on tests given at the end of the school year. To increase continuity in the data used for RTI decisions, screening batteries should include the measures with predictive validity for the primary outcome (e.g., state standards test) and those that will be used to monitor student progress and evaluate intervention effects. Doing so will link the decision-making throughout the course of intervention.

Evaluating Intervention Effects Screening is the critical first step in RTI assessment since it establishes who is potentially at risk for academic difficulty. RTI models differ in whether students initially screened are placed in a tier 2 intervention, or whether those students are first progress-monitored in tier 1 to confirm the initial screening results. In the latter case, monitoring student progress in tier 1 actually becomes part of initial screening. In either case, the process eventually leads to some students being placed in a tier 2 intervention. It is in tier 2 that RTI models must evaluate intervention effects. Different options exist as to how this is to be done.

The simplest option for evaluating tier 2 intervention success has been to establish benchmark scores for fall, winter, and spring (Marston 2012), and then to test those students in winter and spring to determine at those points in time whether students have attained the benchmark scores. If students are only tested at these two points in time (winter and spring), then intervention effectiveness is determined by the percentage of students who meet the benchmark score at those points in time. In this case, the intervention is considered to be effective for those students who meet benchmark in the winter or spring, and they can either be returned to tier 1 or can continue in tier 2. The intervention is considered ineffective for students who do not pass the benchmark and they are either continued in tier 2 with some change in program or considered for more intensive intervention in tier 3.

While testing students in winter and spring to determine who are succeeding might seem like an appropriate approach to evaluating interventions, it suffers from two important problems. First, the benchmarks might be unreasonably high for many students given their initial level of performance at screening. Those students might be growing quite well in the intervention, but their growth rates will not be evident if all we learn is that they do not meet the benchmark. Our evaluation of intervention success in these cases is necessarily based on insufficient evidence and might lead to erroneous conclusions. At the very least, a trajectory can be established for students by connecting their fall performance score with their winter score, or winter score with spring score and using trend to supplement students' score relative to the benchmark. The risk in using this approach is that score trends are based on only beginning and ending data points that possess some degree of error, so the growth rates actually might be steeper or flatter than those obtained by connecting the two raw score points.

A second problem with testing only in winter and spring is that it limits formative decisions to two per year—once in the winter and once in the spring—and the second decision is too late. A truly responsive approach to intervention should attempt to increase the frequency of opportunities to make data-based decisions. This, of course, requires more frequent monitoring of student performance ranging from weekly to monthly. As pointed out earlier, technical problems exist in using progress-monitoring data to establish trends; however, data also exist showing that the stability of student progress-monitoring data can be improved by combining several scores (Ardoin and Christ 2009). It is important for students in tier 2 that we try to generate data that will allow us to decide if we should continue in tier 2, because the intervention is succeeding, or if they are "flatlining" (e.g., not making progress) in the current tier 2 intervention. If the student is not succeeding in the current tier 2 intervention, then either a new tier 2 intervention should be tried or the student should be considered for more intensive intervention at tier 3. It is preferable, of course, to attempt multiple tier 2 interventions rather than moving students to tier 3, which increases their distance from the classroom program and will be much more resource intensive (Fuchs et al. 2010). Eventually, however, our data might indicate that tier 3 intervention is necessary. Indeed, research on this issue reveals that as many as 12 % of the students in general education classrooms (McMaster et al. 2005), 30 % of "at-risk" students (Al Otaiba and Fuchs 2002), and 50 % of students with disabilities (Fuchs et al. 2002) are likely to be nonresponsive to the types of interventions that can be implemented at tier 1 and tier 2.

A strong case can be made that as students require more intensive intervention, the need for more frequent progress monitoring is necessary. The situation is analogous to what happens in health care as people require more intensive treatment for illness. Wellness checks occur only periodically, but when illness strikes, the frequency of vital signs monitoring increases. Eventually, for those who are seriously ill and require intensive care, monitoring is continuous. Students in tier 3 have arrived there because they have not responded positively to interventions at tier 1 and tier 2. The reality of a tier 3 placement is that we now have used the evidence-based interventions we have been able to implement for groups of students at tiers 1 and 2, and still have some students who are not progressing satisfactorily. When evidence-based interventions have not been successful at tiers 1 and 2, we should then, typically, attempt to create more intense interventions by using one-to-one or individually tailored instruction at tier 3 in an attempt to produce more positive outcomes. The question now becomes how to implement and evaluate interventions at tier 3.

Experimental Teaching in Tier 3. Intervention at tier 3 must, necessarily, be characterized by regular and frequent monitoring of student performance and systematic testing of multiple alternative interventions in an effort to find approaches that work for individual students. Typically, all group-based interventions will have been exhausted at tiers 1 and 2. At tier 3, the teacher becomes an applied researcher experimenting with creative, previously untried interventions. The research-based nature of this approach has been aptly referred to as "experimental teaching" (Casey et al. 1988; Fuchs et al. 2010) because teachers in tier 3 need to engage in empirical tests of unique intervention hypotheses designed for individual students. Experimental teaching is illustrated in Fig. 1. The figure shows one student's progression from tier 1 through to tier 3. As can be seen in the figure, at the outset the student is performing well below his peers in reading since he is reading about one third as many words correctly from text as they do in 1 min. Notably, the error scores are quite low. The tier 2 placement is indicated by a hatched vertical line. At that point, a year-end goal is established and a desired growth rate is depicted by connecting the student's performance level at the beginning of tier 2 placement to the goal. Throughout, progress is monitored weekly using a CBM approach where the student reads aloud from three text passages of equivalent difficulty. The median score from those readings is plotted on the graph showing the performance level for that week. The median of three scores is used as the progress datum to increase the stability of that weekly score.

Two different interventions were tried in tier 2 (IV #1 and IV #2). While the progress data seem to indicate that IV #1 was going to be effective, the initial increase in performance was not maintained. As a result, a second intervention, IV #2, was implemented. Still, the scores did not increase satisfactorily. Given the apparently unsatisfactory response to the interventions in tier 2, the student was placed in tier 3 (second hatched line), and IV #3 was implemented. This initial tier 3 intervention had an apparent strong effect to increase the student's rate of growth for about 1 month, at which point the teacher became concerned that the student was leveling off. For that reason, she "tweaked" IV #3 to produce a second tier 3 intervention. The effect seemed to be to produce more variable performance; however, by the end of the school year, the student's scores were very close to that specified in the annual goal.

Successful experimental teaching almost certainly requires creative teachers who are open-minded about what interventions they might try and interested in creative problem-solving. Commitment to one or two approaches to helping nonresponsive students to learn is not likely to be effective. Two findings in the study by Casey et al. (1988) are especially relevant here. The first is that, when monitored across the duration of the project, the teachers using the experimental teaching approach actually became less certain about the strength of the interventions they initially identified as most effective. The researchers attributed this to the fact that the teachers had been testing the effects of those favored interventions over the course of a school year and that they had not seen the growth for all students who they had associated with those interventions. That experience probably had produced suspended judgment about intervention effectiveness on the part of the teachers. Since being open-minded is one of the most important characteristics of a scientist and an experimental teacher, this increased uncertainty was viewed as a desirable outcome. A second interesting finding of the study is that by the end of the yearlong experimental teaching experience, the teachers were able to generate more alternative intervention ideas than at the beginning of the project. This is an ideal outcome for tier 3 teachers, of course, because success at that level in RTI certainly requires not only suspended judgment but also the ability to generate novel interventions and test them in the progress data.

Figure 2 provides a clear picture of why teachers at tier 3 must collect data and evaluate the impact of their instruction for individual students. The figure is an example of experimental teaching with two students across a large segment of the school year. As can be seen in the figure, the vertical intervention lines occur at the same point on the graphs for both students. These lines, of course, indicate that the teacher was creatively changing the intervention for both students at the same time. In addition, the teacher was making the same intervention changes for both students at the same time.

The results of the series of interventions are very clear. The student in the top graph responded well to the successive intervention changes made by the teacher, while the student in the bottom graph very nearly flatlined during the same period. An interesting sidelight in these graphs is that, in almost every case, the change in intervention actually seemed to have a negative impact on the student in the bottom graph. Just prior to each change that student's rate of growth began to increase, but after the change in intervention that student's performance dropped back. The picture reminds us of the tragic figure of Sisyphus in Greek mythology who was consigned to spend eternity pushing a boulder up a hill and have it roll back down on him before he reached the top.

The teacher working with the two students in Fig. 2 was implementing an experimental teaching approach in all respects except one. While he was responsive to the progress-monitoring data for the one student, he was not responsive to the other. In large part, this occurred because the teacher was teaching both students at the same time. One has to wonder whether the sessions with those two students might have been organized so that the instruction could have been differentiated. This is not an easy task, of course, but if one has to succeed with previously nonresponsive students, individually organized interventions might be the only solution. If one is going to be successful in conducting effective RTI programs, it will be well to remember that that one cannot simply implement evidence-based interventions and assume that they will be effective. A paraphrase of one of Abraham Lincoln's sayings provides a good summary of this point:

> Interventions might exist that are effective with all students for a short time, and interventions exist that are effective with some students for an extended time, but no interventions are effective with all of the students all of the time.

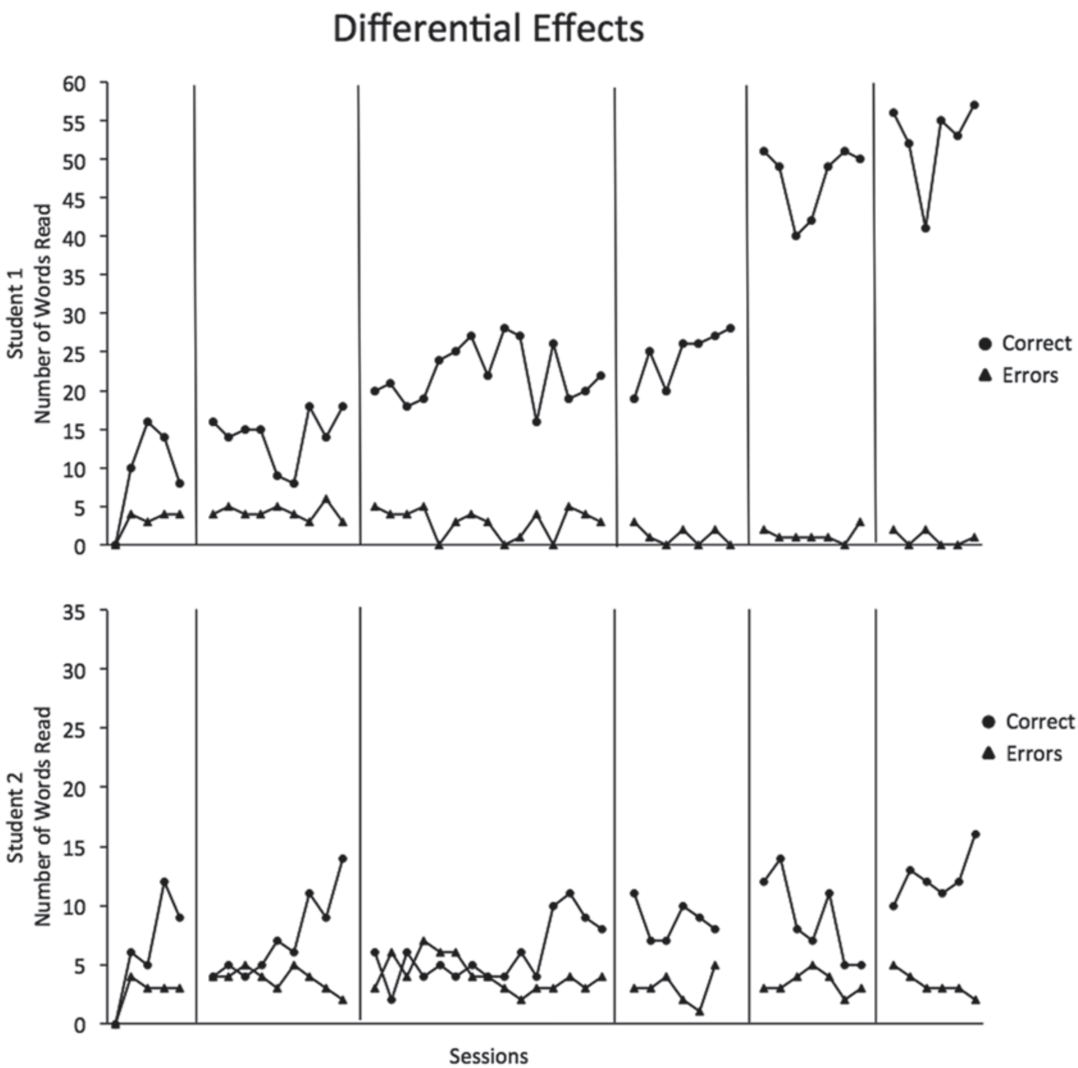


Fig. 2 Differential effects of the same interventions with two students

Conclusion

At the outset of this chapter, the I provided a personal perspective on how I came to believe a data-based approach to making instructional programming decisions was both necessary and helpful. That experience led me to develop a framework for thinking about making data-based decisions that rested on five assumptions about the purpose of interventions and what was required for those interventions to be effective. In the remainder of the chapter, I considered current developments in RTI from the frame provided by those early assumptions. It seems clear to me that the requirements driven by those early assumptions are applicable to current efforts to make data-based intervention decisions. Interventions in RTI need to be evaluated, agreed upon vital signs of student progress are still required, and well-trained professionals are necessary for RTI decisions to be data based and effective. Implications of data-based decision-making for practice are briefly summarized in Table 1.

Consideration of current efforts to data-based decisions in RTI also led me to raise questions and issue a few caveats for those professionals involved in developing RTI programs. I cannot help but wonder how practically feasible it is to monitor performance frequently across large numbers of students within a school. If the reality is that students will only be tested at the beginning, middle, and end of the school year, I wonder how responsive RTI programs are to individual differences in rates of growth. I wonder, too, whether typical progress-monitoring data are the most reliable and valid for making program evaluation and placement decisions or whether more extensive test batteries are better suited for those decisions. A great deal of research needs yet to be done on cut scores for different assessment devices to assure that appropriate students are identified for service and precious resources are not misallocated. As cut scores are identified, consideration must be given to the degree that identification decisions will be actuarial and the extent that clinical judgment will be incorporated.

Despite the fact that many questions remain to be addressed, the future of data-based decision-making in the schools appears to be very promising. Considerable research, development, and field testing of tools used to make data decisions have been conducted. The feasibility of taking a data-based approach seems now to be beyond question. Computer-based technology has begun to contribute to both the management of student progress data and to aid in making data-based decisions. None of this was available when the original work on DBPM was conceptualized. None of us involved in the early development

Table 1 Implications of data-based decision-making for practice

1.	Establish common goals and the data that will be used by all to determine whether the goals are being met
2.	Choose long-range goals on which progress can be measured for at least an entire school year so that interventions can be evaluated using the data
3.	Treat interventions as hypotheses whose effects will be revealed in the data and be prepared to try alternatives when interventions are not leading to goal attainment
4.	Continually work at improving the reliability and validity of the data and the criteria you are using to decide whether students should continue in their current intervention levels or should be moved to different levels
5.	Create regular in-service training procedures to assure that all those collecting and using data to make decisions understand how to collect the data, why the data are being collected, how to interpret the data, and how to make the decisions
6.	Increase the frequency with which student progress is measured and the responsiveness of the intervention system as the students move to more intense levels of intervention
7.	Recognize that even evidence-based interventions do not work for every student and design your program in such a way as to enable teachers to find or create and test alternative interventions when evidence-based interventions have not been effective

imagined that school-wide applications now underway would be possible. The content of this handbook is a tribute to the creativity and sustained effort of many who believe that struggling students deserve a data-based approach to making intervention decisions designed to help them succeed.

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy instruction: A review of the literature. *Remedial and Special Education, 23*(5), 300–316.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Estimates of standard error when monitoring progress using alternate passage sets. *School Psychology Review, 38,* 266–283.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbell, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18.

Behn, R. D. (2009). The data don't speak for themselves. Bob Behn's Public Management Report, 6(7). Cambridge: Harvard University.

Brandt, S. (2012, November 26). Minnesota educators push to end high-stakes math exam. *Star Tribune*, (pp. A1, A5).

Casey, A., Deno, S. L., Marston, D. B., & Skiba, R. (1988). Experimental teaching: Changing beliefs about effective instructional practices. *Teacher Education and Special Education, 11*(3), 123–131.

Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology, 47,* 55–75.

Christ, T. J., & Coolong-Chaffin, M. (2007). Interpretations of curriculum-based measurement outcomes: Standard error and confidence intervals. *School Psychology Forum: Research in Practice, 1*(2), 75–86.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98,* 394–409.

Davis, G. N., Lindo, E. J., & Compton, D. (2007). Children at-risk for reading failure: Constructing an early screening measure. *Teaching Exceptional Children, 39*(5), 32–39.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219–232.

Deno, S. L. (1995). School psychologist as problem solver. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (pp. 471–484). Washington, DC: National Association of School Psychologists.

Deno, S. L. (2005). Problem-solving assessment with Curriculum-based Measurement (CBM). In R. Chidsey-Brown (Ed.), *Problem-solving based assessment for educational intervention*. New York: Guilford.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston: Council for Exceptional Children.

Deno, S. L., Wesson, C. L., Mirkin, P. K., Maruyama, G., Skiba, R., King, R. P., & Sevcik, B. (1988). A causal analysis of the relationships among ongoing curriculum-based measurement and evaluation, the structure of instruction, and student achievement. *Journal of Special Education, 22*(3), 330–343.

Espin, C. A., McMaster, K. L., Rose, S., & Wayman, M. M. (2012). *A measure of success: The influence of curriculum-based measurement on education*. Minneapolis: University of Minnesota Press.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–501.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984) The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*(2), 449–460.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989a). Computers and curriculum-based measurement: Effects of teacher feedback systems. *School Psychology Review, 18,* 112–125.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989b). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10*(2), 43–52.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., & Braun, M. (2002). Preliminary evidence on the importance of reading programs for kindergartners with disabilities in mainstream classrooms. *Exceptional Children, 68*(3), 295–311.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71,* 7–21.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (2005). Using technology to facilitate and enhance curriculum based measurement. In K. Higgins, R. Boone, & D. Edyburn (Eds.), *The handbook of special education technology research and practice* (pp. 663–681). Whitefish Bay: Knowledge by Design, Inc.

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new general education continuum of placements and services. *Exceptional Children, 76,* 301–323.

Grove, W. M., & Lloyd, M. (2006). Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology, 15*(2) 192–194.

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson, & D.

Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–149). Mahwah: Erlbaum.

Jenkins, J. R., Zumeta, R., Dupree, O., & Johnson, K. (2005). Measuring gains in reading ability with passage reading fluency. *Learning Disabilities Research and Practice, 20*(4), 245–253.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review, 36,* 582–599.

Jenkins, J. R., Graff, J. J., & Miglioretti, D. (2009). Validity of reading growth estimates. *Exceptional Children, 75*(2), 151–164.

Lembke, E., McMaster, K. L., & Stecker, P. M. (2012). Technological applications of Curriculum-Based Measurement in elementary settings. In C. A. Espin, K. L. McMaster, S. Rose, & M. M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 125–138). Minneapolis: University of Minnesota Press.

Marston, D. B. (2012). School- and district-wide implementation of Curriculum-Based Measurement in the Minneapolis Public Schools. In C. A. Espin, K. L. McMaster, S. Rose, & M. M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 59–78). Minneapolis: University of Minnesota Press.

Marston, D. B., Deno, S. L., & Fuchs, L. S. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-based measures. *Diagnostique, 11*(2), 77–90.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to non-responders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*(4), 445–463.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed, pp. 13–103). New York: Macmillan.

National Center on Response to Intervention. (2012). http://www.rti4success.org/.

Reynolds, M. C., & J. W. Birch. (1982). *Teaching exceptional children in all America's schools*. Reston: The Council for Exceptional Children.

Skiba, R., & Deno, S. L. (1984). Review of single subject research in special education and single-case research designs: Methods for clinical and applied settings. *Exceptional Children, 51*(1), 81.

Skiba, R. J., Deno, S. L., Marston, D., & Wesson, C. (1986). Characteristics of time-series data collected through curriculum-based reading measurement. *Diagnostique, 12,* 3–15.

Skiba, R. J., Deno, S. L., Marston, D., & Casey, A. (1989). Effects of trend estimation and subject familiarity on practitioners' judgments of intervention effectiveness. *Journal of Special Education, 22*(4) 433–446.

Speece, D. L. (2012). Curriculum-based Measurement progress monitoring and the health of general education. In C. A. Espin, K. L. McMaster, S. Rose, & M. M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 59–78). Minneapolis: University of Minnesota Press.

Stiggins, R. J. (1993). Teacher training in assessment: Overcoming the neglect. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills*. Lincoln: Buros Institute of Mental Measurement University of Nebraska.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R. & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41,* 85–120.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus: Merrill.

Applied Behavior Analysis: A Foundation for Response to Intervention

Scott P. Ardoin, Liliana Wagner and Kathryn E. Bangs

Implementation of a *response to intervention* (RTI) model requires schools to make both conceptual and procedural modifications that are essential to its success and will lead to improvements in the education and behavior of all students within a building. Schools must modify their identification process from one which identifies children based upon discrepancies in the constructs of intelligence and achievement to one that identifies students who are at risk of later disability identification. RTI is a prevention-based model that requires students to receive early intervention based upon their instructional and behavioral needs. Waiting for students to fail before beginning to provide them with intervention results in too many students being identified as needing intervention when the possibility of remediation is slim (President's Commission on Excellence in Special Education 2002). A second change required by schools in order to ensure the success of RTI is a change in assessment procedures. Schools must not rely so heavily on evaluating behaviors believed to represent internal states/constructs and instead directly evaluate the behaviors and skills students must possess to be socially and academically successful. Measuring behaviors that students need to succeed, as opposed to measuring constructs, will provide schools with data that can directly inform instruction. Finally, implementation of an RTI model requires that schools do not blame children for their behavior or poor academic performance, and instead determine what aspects of the instructional environment must be manipulated to maximize student achievement. In essence, the implementation of RTI within schools requires a shift from providing services to students based upon a nomothetic traditional assessment framework to assessing student needs based upon an idiographic applied behavioral analysis (ABA) framework.

Traditional Assessment The goal of traditional assessment is to determine the precise problem and measure inferred states from indirect observations (Galassi and Perot 1992; Hayes et al. 1986; Tkachuk et al. 2003). These inferred states are assumed to be stable, intraorganismic variables (Hayes et al. 1986; Nelson and Hayes 1979a, 1979b). That is, the variables do not change based on the situation, rather they are a reflection of the person's cognitive constructs, mental disorders, and/or their personality emphasizing personology constructs (Hayes et al. 1986; Nelson and Hayes 1979a; Tkachuk et al. 2003). Hence, behavior is seen to be a sign or sample of these constructs (Hayes et al. 1986; Nelson and Hayes 1979b). Behavior is not viewed as having controlling variables outside of the individual; therefore, these variables are largely ignored (Hayes et al. 1986). Since traditional assessment is based on stable constructs, the focus of

S. P. Ardoin (✉) · L. Wagner · K. E. Bangs
Department of Educational Psychology,
University of Georgia, 323 Aderhold Hall,
30602 Athens, Georgia
e-mail: spardoin@uga.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_3

assessment is to classify and diagnose individuals, therefore, emphasizing group experimental designs (Tkachuk et al. 2003) and the comparison of individual performance to normative data (Hayes et al. 1986).

The IQ-achievement discrepancy model, a traditional assessment approach, has served as the approach for determining students' special education eligibility since the 1970s (Fletcher et al. 2004). Under the discrepancy model, students are identified as needing special education if a discrepancy exist between the constructs of intelligence (believed to represent learning potential) and academic achievement (Sparks and Lovett 2009). Other constructs (e.g., inattention, depression, anxiety) are also measured to evaluate whether the discrepancy between their intelligence and achievement is due to a learning disability or the manifestation of other internal states. To measure these constructs, caregivers (e.g., teachers and parents) as well as the students themselves, might be asked to complete behavioral rating scales on the frequency with which a student engages in various behaviors (e.g., day dreaming, interrupting others, fidgeting, somatic complaints). Students with attention problems might have higher ratings on questions pertaining to concentration, perceptions of being off-task, and boredom in school, whereas students with anxiety would have higher ratings on questions about fearing people or situations, being self-conscious, and being nervous (Achenbach 1991; Reynolds and Kamphaus 2006).

Basing students' educational needs primarily on their internal states, as determined by samples of their behavior believed to represent constructs, ignores the impact of the learning environment on behavior. Ignoring the impact of the environment results in students being blamed for their failure and, ultimately, labeling some students with a disability who were, in fact, disabled by their learning environment. Choosing to focus on problems within the child also limits the availability of information used for developing effective interventions (President's Commission on Excellence in Special Education 2002).

Applied Behavior Analysis The discipline of ABA provides a foundation from which the principles and procedures for RTI were developed. ABA is the science of solving socially important problems by evaluating how the environment impacts behavior (Gresham 2004). Emphases of ABA include (a) measuring individuals' behavior as opposed to their mental states, (b) continuity between those behaviors which are observable, and those events private to the individual, (c) predicting and controlling individuals' behavior as opposed to the behavior or mental states of groups, and (d) an understanding of the environmental causes of behavior (Fisher et al. 2011). The primary focus of ABA is on observable behaviors (e.g., talking, academic engagement, biting, crying, fidgeting, etc.) and measuring the ways in which the environment (e.g., quality of instruction and teacher attention) influences the dimensions of those behaviors (Cooper et al. 2007). In ABA, the amount (frequency, intensity, or duration) a person engages in target behaviors of interest is compared to his/her history of engaging in the target behavior under the same or different environmental conditions. Environmental variables are measured and manipulated to evaluate how changes in the environment might alter the amount a person engages in the behavior.

With ABA as a foundation, RTI models focus on whether the general education environment (tier 1) can be expected to produce adequate learning and the environmental changes necessary for producing significant gains for the target student. Assessing the skills and behavior of all students within a school through multiple universal screenings across the academic year provides RTI teams with information regarding the general effectiveness of tier 1 instruction. It allows the academic performance and classroom behavior of individual students to be compared to other students receiving similar tier 1 instruction and behavior management. Manipulation of the existing environment through the implementation of increasingly intensive interventions during tier 2 instruction provides information to RTI teams regarding the level of modifications necessary to produce adequate learning and behavior. Students who require environmental manipulations that extend beyond that which a general education classroom can provide are in need of special education services (tier 3) (Fletcher et al. 2005; Marston 2005; Shinn 2007). Given ABA's focus on behavior, the environment, and measuring the impact of changes to the environment on the behavior of individuals, it provides a framework that RTI teams can reference when (a) identifying target behaviors, (b) selecting and developing interventions, and (c) measuring intervention effects (Martens and Ardoin 2010) (see Table 1.)

Selecting Target Behaviors

An essential feature of ABA as an *applied* science is the targeting of behaviors for change that are immediately important to the individual and society (Cooper 1982). At face value, selecting target behaviors for change within a school setting that are socially significant and will result in immediate changes in student behavior may sound simple. The task, however, requires knowledge of appropriate replacement behaviors, an understanding of the hierarchy of skills necessary to perform complex tasks

Table 1 Applied behavior analysis and RTI: Implications for practice

Selecting target behavior: Target behaviors should be selected that are immediately important to the individual and society	
	Behavior: Although the purpose of intervention is generally to eliminate one or more inappropriate behaviors, it is best to identify appropriate replacement behaviors. Replacement behaviors should
	Allow the student to access to the same consequence that inappropriate behavior leads to as otherwise the student will still have a need to engage in the inappropriate behavior
	Be as easy for the student to engage in as the inappropriate target behavior
	Lead to reinforcement at least as readily as the inappropriate behavior
	Academic: Although the eventual goal may be to have the student perform complex skills at a level commensurate with peers, it is essential that students have the prerequisite skills necessary to perform the complex behaviors. Identify what target prerequisite skills the student does and does not possess and then target those skills for which the student needs greater accuracy and fluency
Selecting and developing interventions: To increase the probability of selecting an appropriate intervention, schools should experimentally test their hypotheses as to why a student is struggling behaviorally or and/or academically	
	Behavior: Although schools frequently collect indirect and descriptive assessment data through functional behavioral assessments, the hypotheses generated from these data are rarely tested. Modified functional analyses procedures that are appropriate for implementation within schools could enable schools to tests the hypotheses that they generate and increase the probability that they develop a functionally relevant and effective intervention
	Academic: Functional behavioral assessments are rarely thought of in relationship to determining why a student's academic performance is not adequate. Daly and colleagues (1997, 2002) have developed hypothesized functions of students' academic difficulties as well as brief experimental analyses procedures for testing the generated hypotheses

Table 1 (continued)

Selecting target behavior: Target behaviors should be selected that are immediately important to the individual and society	
Measuring intervention effectiveness: It is essential that the behaviors/academic skills measured to evaluate intervention effectiveness are likely to change immediately as a function of the intervention being implemented	
	Behavior: Single subject designs can be used to empirically demonstrate that an intervention is the cause of improvements in observed behavior. Demonstration of a functional relationship requires evidence of (a) *prediction:* behavior changes in the predicted direction with the implementation of intervention, (b) *verification:* when intervention is withdrawn, behavior worsens or the behaviors on which intervention was implemented do not change when intervention is implemented on other target behaviors, and (c) *replication:* the effects of intervention are replicated when re-implemented with on the same behavior or when implemented on new behavior/participants/settings on which intervention was not previously implemented
	Academic: RTI teams generally collect data intended to assess students' response to intervention. Data also need to be collected that allow for the determination of whether an intervention is simply effective
	Data need to be collected that measure the effect of intervention on the skills being targeted
	Measuring skills targeted through intervention may require that the variable(s) being measured change across the period of intervention implementation, but measuring targeted skills as opposed to generalization will provide a more sensitive measure to intervention effects
	Although measuring targeted skills may require RTI teams to collect (a) data that inform intervention effectiveness and (b) data that inform responsiveness to intervention, it should allow for teams to determine more quickly that an intervention is unlikely to lead to a student adequately responding to the intervention. More resources may be required to collect data, but fewer resources will be wasted on implementation of ineffective interventions

(e.g., reading with comprehension), and knowing how to assess the behaviors targeted by intervention (Noell et al. 2011)

Classroom Behavior When developing interventions for students who are engaging in unacceptable levels of inappropriate classroom behavior, research within ABA has demonstrated the importance of identifying appropriate alternatives that could potentially replace the inappropriate behavior (Carr and Durand 1985; Cooper et al. 2007). Although punishing inappropriate behavior will likely result in an immediate decrease in punished behaviors, failure to teach an appropriate replacement behavior may result in the student simply replacing the inappropriate behavior with an equally problematic alternative. For this reason, education professionals implementing an RTI framework must determine why the student is engaging in inappropriate behavior (i.e., what is reinforcing the inappropriate behavior), what appropriate behavior(s) might enable the student to access the same reinforcer(s), and what appropriate replacement behavior(s) the student is capable of engaging in. RTI teams should refer to the large base of ABA research using functional assessment procedures to identify the factors potentially reinforcing inappropriate behavior, and the literature on differential reinforcement for guidance on selecting and increasing appropriate alternatives (Geiger et al. 2010; LeGray et al. 2010; Miller and Lee 2013).

Academic Performance Selecting a target behavior for students identified as needing tier 2 intervention due to poor academic performance might seem obvious, as the goal is to get the student to accurately perform the academic tasks related to his/her area of weakness. In fact, it may be tempting to simply target the behavior(s) measured through universal screening procedures which were used to identify the student as needing intervention. However, by definition the students identified through a universal screening as needing tier 2 intervention are not performing at the level of their peers, and are likely in need of remediation of prerequisite skills to adequately perform those behaviors at which their peers are succeeding. Thus, the tendency to select academic behaviors within the child's grade level curriculum is unlikely to be an appropriate target for intervention (Brown-Chidsey and Steege 2005; Daly et al. 2007). For instance, many schools employ curriculum-based measurement procedures in reading (CBM-R) to identify which of their students might need supplemental reading instruction. CBM-R measures students' oral reading rate with accuracy (Ardoin et al. 2013; Mellard et al. 2009). Second grade is recognized as a grade in which fluency greatly increases across the academic year. It would, therefore, seem socially appropriate to provide second grade students struggling in reading with an intervention addressing reading fluency. A lower-performing second-grade student may not, however, have the prerequisite skills (e.g., letter recognition, knowledge of letter sounds, phonological awareness) necessary to benefit from an oral reading fluency-based intervention. The intervention therefore would not likely result in the student making adequate progress, not due to the student needing a more intensive intervention, but due to the student needing an intervention targeting his/her skill needs. Thus, although oral reading fluency might seem to be the most socially significant behavior to target, such a decision would be unlikely to elevate the student's skills to a level comparable to his/her peers.

Selecting the proper academic skills to target for intervention requires knowledge of the prerequisite skills necessary to perform the complex tasks that make up grade level curriculum goals. Research in ABA highlights the importance of creating task analyses for complex, multistep activities, and then determining which of the skills the student has and has not mastered (Noell et al. 2011). For instance, Parker and Kamps (2011) used task analyses, in combination with self-monitoring and social scripts, to teach functional life skills and increase verbal interactions with peers in two children with autism. Three activities were selected for the intervention (playing board games, cooking, and dining in a restaurant) and a task analysis was developed for each. The task analyses were written and displayed on a laminated sheet of paper, listing each step necessary for completing the activity in the order it needed to be completed. The number of steps for each activity ranged from 8 to 22, and, in addition to performing the more functional components of the task, the steps included prompts for initiating conversation with peers. Participants were taught how to use the task analysis, using a combination of verbal and physical prompts, in the settings in which the intervention sessions would take place. Data were collected on each student's task completion, activity engagement, and peer directed verbalizations. Once students' proficiency in task completion improved, prompts were faded in order to decrease reliance on the prompts and increase the probability of the student generalizing the behaviors to other settings. Similar procedures can be employed for teaching students how to complete complex math problems or write well-constructed narratives (Alter et al. 2011; Hudson et al. 2013; Noell et al. 2011).

Selecting and Developing Interventions

A primary objective of ABA is to identify the motivation behind behavior, and use that knowledge to create individualized interventions that work to reduce problem behavior and increase positive replacement behaviors (Gresham 2004). This is synchronous with the goal (within an RTI framework) of determining whether a student's problem behavior or poor academic performance is a function of the environment in which they are being instructed, or the result of a disability. Universal screening data that indicates other students are responding to tier 1 instruction provides evidence that general education instruction is adequate. If interventions are properly selected, students' response to increasingly intensive levels of intervention through tier 2 instruction should provide RTI teams with information regarding whether (a) a student failed to learn essential prerequisite skills that if learned would allow the student to make adequate academic progress and (b) the level of modifications necessary to tier 1 instruction for the student to make adequate rates of academic progress (McDougal et al. 2010). The likelihood that tier 2 will provide RTI teams with accurate information regarding a student's instructional needs is largely based upon whether proper analyses are conducted to determine the function of students' poor classroom behavior or academic performance.

Classroom Behavior Research in ABA suggests that development of an intervention based upon the cause of the student's problem behavior is likely to lead to greater improvements in behavior with a less intense intervention (Daly et al. 1997; Iwata et al. 1994; Vollmer and Iwata 1992). Determining the cause of problem behavior is often accomplished through a functional behavior assessment (FBA), a process of analyzing environmental conditions and collecting data on patterns of behavior to establish a hypothesis of the function (Cooper et al. 2007; Solnick and Ardoin 2010). The term FBA is familiar to many educators, especially those involved in special education. Amendments to the Individuals with Disabilities Education Act (IDEA) in 1997 mandated that schools conduct an FBA, and implement a behavioral intervention plan based on the behavioral function, if a student's aberrant behavior is determined to be a result of his/her disability (Individuals with Disability Education Act Amendments of 1997 [IDEA] 1997).

FBAs involve a process of gathering information on the function of behavior through indirect assessments (e.g., structured interviews, rating scales), descriptive assessments (e.g., direct observations in the child's typical environment), and functional analyses. The data collected from these procedures are used to form and test hypotheses about the motivation behind a child's behavior (Cooper et al. 2007). Once the function of problem behavior is identified, targeted interventions can be put in place to change the reinforcement contingencies maintaining the problem behavior. Most behavioral interventions developed through the widely used school-wide behavioral management system known as *positive behavioral intervention and support* (Positive Behavioral Interventions and Supports (PBIS) 2013; Sugai et al. 2000) begin with an FBA (Carr and Sidener 2002), and use the information collected to create interventions teaching the student more appropriate, effective, and efficient ways of accessing reinforcement than engaging in the inappropriate behavior (Carr and Sidener 2002; Sugai and Horner 2002). The basis for these assessments is derived from the ABA research literature (Iwata et al. 1994; Vollmer and Iwata 1992).

IDEA requirements for FBAs in schools were largely based upon research conducted with students engaging in high rates of severe behavior, and studies in which a component of functional assessment, functional analysis, was implemented (Drasgow and Yell 2001; Individuals with Disability Education Act Amendments of 1997 [IDEA] 1997, 2004). Functional analysis is the only component of an FBA that involves the direct testing of hypotheses on behavioral function by performing systematic environmental manipulations and examining which of the test conditions elicits the highest rate of problem behavior. In a school setting, the conditions typically assessed involve attention (positive reinforcement), escape from demands (negative reinforcement) and play (automatic or sensory reinforcement; Solnick and Ardoin 2010). For example, in the typical functional analysis escape condition, students are issued a demand and allowed to momentarily escape from the task demands when they engage in problem behavior. If the function of student behavior is escape, then the student is more likely to escape in this condition. Thus, the condition that results in the highest rate of inappropriate behavior is hypothesized to be the function of that behavior. Although it might be considered problematic to intentionally induce problem behavior in this manner, with severe problem behavior (e.g., head banging), the benefits of creating an effective intervention may outweigh the risk of the student engaging in inappropriate behavior during assessment sessions (Iwata et al. 1994).

Functional analyses are difficult to implement in a school setting due to the challenges of controlling the classroom environment (e.g., peer attention) and the fact that functional analysis conditions are meant to elicit problem behavior, which may endanger the student, peers, and teachers (Bloom et al. 2011; Solnick and Ardoin 2010). Several studies have, however, demonstrated some success in employing modified functional analysis procedures in the child's natural environment. Bloom et al. (2011) substantially reduced the length of sessions to only 2 min per test condition and implemented the test conditions within the context of naturally occurring classroom activities. For example, attention and tangible conditions were conducted during free play, and the demand conditions occurred during instructional time, when it was appropriate for the teacher to be issuing demands. The authors reported that results from the trial-based functional analyses matched those of a standard functional analyses conducted for comparison purposes in 60 % of the cases. In another modification to traditional functional analyses procedures, researchers have examined participants' latency to problem behavior under different levels of aversive demands and teacher attention (Call et al. 2009; Hine and Ardoin 2011). Such procedures allow for analyses to be conducted with fewer occasions of inappropriate behavior and allow for inappropriate behavior to be appropriately dealt with when it occurs.

Functional analysis procedures have also been successfully implemented by classroom teachers with typically developing children at risk for reading failure, displaying high levels of off-task and disruptive behavior (Shumate and Wills 2010). All functional analysis sessions in this study were conducted during regularly scheduled classroom activities, with escape and attention conditions occurring during small group reading instruction and the control condition taking place in a separate play area. Three sessions were presented each day (one of each condition), until the data showed a clear pattern of responding. As teacher attention was determined to be maintaining the aberrant behavior of all participants, an intervention was designed to address this specific function. Teacher attention was withheld for all instances of problem behavior, with immediate attention and praise-delivered contingent on desirable alternative behaviors (e.g., hand raising). The results from this Shumate and Wills (2010) indicate that the function-based intervention was successful in both decreasing the participants' off-task and disruptive behaviors and increasing appropriate alternatives, further validating the use of function-based assessments in the role of identifying the variables maintaining problem behavior.

Academic Performance The FBA literature might initially not seem to generalize to students difficulties with academic skills. However, the benefits of systematically altering the stimuli students are exposed in order to evaluate the causes of their academic difficulties remain, whereas, in the FBA literature, problem behavior is believed to be a function of attention, escape, tangibles (e.g., access to toys or food), or automatic reinforcement (Cooper et al. 2007). Daly et al. (1997) argued a student's failure to perform academically is a function of his/her (a) lack of motivation, (b) insufficient opportunities to practice the skill, (c) not previously having sufficient assistance/instruction/modeling in how to perform the task, (d) not having been asked to perform the task in that manner previously, and (e) simply not having the prerequisite skills necessary to perform the task. In line with this theory, researchers have explored brief experimental analysis (BEA) procedures, which include the implementation of test conditions that involved implementing interventions matching the aforementioned function of poor academic performance. Daly et al. (2002) conducted a BEA of reading during which elementary students were exposed to interventions of increasing complexity that were associated with the functions of poor academic performance identified by Daly et al. (1997). The intervention that produced the highest level of student reading accuracy and fluency for each student was hypothesized to be associated with the function of each student's poor reading fluency.

The interventions tested within BEA of academic performance are largely based upon Haring and Eaton's (1978) *instructional hierarchy,* a systematic framework for providing students with prompts to promote correct academic responding, and consequences to encourage future correct responding. Haring and Eaton (1978) described how principles of ABA apply to academic learning. They outlined strategies to move a student from not having the skills or knowledge to respond accurately to stimuli to responding accurately and fluently to that stimuli, and finally, generalizing and, then, adapting knowledge and skills to new instructional materials (Ardoin and Daly 2007). Haring and Eaton's *instructional hierarchy* has been employed across multiple reading and math studies to assess and develop interventions for elementary students (e.g., Ardoin et al. 2007; Cates and Rhymer 2003; Eckert et al. 2002; Martens et al. 1999) and is central to many of the problem-solving models employed by RTI teams in schools (Daly et al. 2005; Goldstein and Martens 2000; Hosp and Ardoin 2008).

Haring and Eaton (1978), as well as supporting studies (e.g., Belfiore et al. 1995; Szadokierski and Burns 2008), suggest that multiple opportunities to respond to stimuli promotes future accurate and fluent responding to the drilled stimuli. For oral reading, this would mean that repeated reading increases accurate and fluent responding to words and word sequences that were drilled. Essentially, multiple opportunities to practice helps to develop strong stimulus control, allowing for accurate and fluent reading of drilled text. After stimulus control is developed to allow for accurate and fluent responding, generalization can be promoted by providing practice opportunities. Practice involves opportunities to respond to learned stimuli when the stimuli are presented across multiple contexts. For oral reading, this would mean providing opportunities to read the same words in different configurations or passages, thus, allowing for further development of stimulus control at the word level and generalization is reinforced (i.e., generalization due to multiple exemplars).

Generalization and Maintenance Another important principle of ABA that must be incorporated into the implementation of interventions within an RTI model is programming for generalization and maintenance of improvements in behavior over time (Mesmer and Duhon 2011). Generalization and maintenance of intervention effects should not be expected as a positive side effect of intervention; rather, they must be programmed into the intervention (Stokes and Baer 1977). Interventions developed within an ABA framework are developed so that improvements in behavior are not only observed within intervention sessions but also generalized to other settings and maintained across time. Steps taken to increase the probability of generalization and maintenance include (a) training behaviors that will be naturally reinforced outside of the intervention setting, (b) conducting the intervention across multiple settings and using multiple exemplars of stimuli that signals to the individual to engage in the appropriate behavior, and (c) training target behaviors to high levels of proficiency, which minimizes the effort required by the student to engage in the behavior and thus increases the probability that the behavior will occur and be reinforced across settings (Ardoin et al. 2007; Mesmer and Duhon 2011). For instance, Ledford et al. (2008) evaluated the effectiveness of a teaching procedure on the acquisition of related and pictorial information by children with autism during sight word instruction conducted in pairs. The authors selected target words from lists provided by caregivers, ensuring the information learned would be relevant, and, thus, more likely to be reinforced in the children's natural environments. In addition, generalization probes were conducted throughout the study in contexts where children could apply the information they were taught in the classroom. For example, a child might be asked to identify a sign reading "Keep Out" while walking past the janitor's closet.

The same steps used within ABA for promoting generalization and maintenance of target behaviors have and should be employed when developing tier 2 behavioral and academic interventions. For instance, many schools employ repeated reading procedures for improving students' reading fluency and comprehension. To ensure the effects of intervention are generalized to the classroom, the materials on which intervention is being provided could be selected from the reading curriculum. Repeated readings could also be provided on directions that are frequently presented on standardized tests and classroom tests, as well as on content area classroom materials (e.g., science, history). If content area materials are too challenging for the student to read, listening passage preview procedures can be provided on the materials, which involve modeling accurate reading and, thereby, preventing students from practicing errors (Eckert et al. 2002). It is also important that interventions be implemented across settings in order to promote students' engagement in appropriate behavior and/or use of newly learned skills across content areas. With preplanning, physical education teachers can easily incorporate basic math calculation and problem-solving instruction into their activities, and content area teachers can assist students in applying targeted reading comprehension and writing skills into their daily instruction. It is, however, essential that all individuals providing instruction/intervention are taught how to implement the procedures as inconsistent implementation can hamper intervention progress.

Measuring Intervention Effects

In addition to providing RTI teams with a foundation for selecting behaviors to target for intervention and developing interventions, ABA provides a model for evaluating intervention effects. Using direct, continuous measurement of individuals' behavior to inform intervention decisions is an essential component of both ABA and RTI (Cooper 1982; Carr and Sidener 2002). Failure to continually evaluate the effects of intervention on classroom behavior and academic performance has the potential to result in the continued implementation of ineffective interventions that can worsen classroom behavior or further increase the discrepancy between the academic achievement of a student and his/her peers.

Selecting Behaviors to Monitor Many RTI teams monitor the effects of students' academic performance using the screening measures employed by their schools for conducting universal screenings (Mellard et al. 2009). Although useful for evaluating generalization effects and for comparing a students' rate of growth to peers, these measures are typically not sufficiently sensitive to measure intervention effects within short periods of time (2–3 weeks). RTI teams should, therefore, monitor the behaviors that are specifically targeted by the intervention being provided to a student (Ardoin et al. 2008). For instance, if an intervention is implemented to improve a student's on-task behavior, on-task behavior collected in the setting and time during which intervention is implemented would be the most appropriate behavior to measure. Although the student's work completion and class grades might also be expected to improve, these two outcomes are not directly targeted by the intervention and are likely to be impacted by environmental variables other than those controlled by the intervention , and, thus, should not serve as primary outcome measures. Likewise, a student's response to an intervention designed to improve vocabulary and comprehension skills should not be evaluated based upon CBM-R (oral reading fluency) progress monitoring data. Even though a students' oral reading fluency would likely improve with gains in vocabulary and comprehension, CBM-R is a general outcome measure on which immediate intervention effects should not be expected (Ardoin et al. 2013). Although there is an emphasis on generalization within the field of ABA, data must be collected on the behaviors directly targeted through intervention as generalization does not occur immediately. Measurement of generalization alone may result in a premature decision to designate an intervention as not producing adequate student gains, when, in fact, the intervention is having positive effects (Ardoin et al. 2008, 2013). Measuring whether a student improves in those skills directly targeted through intervention can provide schools with data at an earlier phase of intervention implementation that will directly inform them of the probability that an intervention will produce inadequate response to instruction (Ardoin 2006).

Evaluating Intervention Effects Neither ABA nor RTI teams typically use inferential statistics to evaluate the effects of intervention on groups of individuals. Instead, visual analyses of data plotted across baseline and intervention conditions are conducted. Despite having a common level of analysis (the individual), as well as common purposes for conducting assessments (i.e., identifying the cause of problem behavior) and implementing interventions (decreasing inappropriate behavior and increasing appropriate behavior, teaching skills), the questions asked within ABA and by RTI teams when evaluating intervention data generally differs. RTI teams are primarily interested in whether or not a student is making adequate progress. When examining classroom behavior, a student's current level or rate of behavior is compared to a desired level of behavior, often taken from the average of other students in the classroom. Evaluating whether a student is making adequate academic gains is generally determined by comparing the student's observed rate of growth to a target rate of growth that is based upon either national or district level normative growth rates (Ardoin et al. 2013).

In ABA, the question is not simply whether a student is responding adequately, but whether changes in behavior are truly due to intervention implementation and if the changes are sufficient enough to positively impact student functioning (Gresham 2004; Roane et al. 2011). In order to address these questions, a student's behavior is not compared to prespecified normative levels or rates of gain. Rather, the level, trend, and variability of intervention data are compared to these same characteristics of baseline data. Baseline data collected as part of an ABA study are intended to provide an understanding of behavior under pre-intervention conditions. Only by understanding pre-intervention behavior can it be known if behavior has changed with the implementation of the intervention. Improvements in behavior from baseline to intervention does not, however, guarantee that observed changes were due to the intervention (Roane et al. 2011).

To demonstrate that the intervention alone is responsible for changes in behavior, single subject design methodology is employed to demonstrate the elements of prediction, verification, and replication. Prediction is demonstrated when behavior changes in the intended therapeutic direction as compared to baseline data. The element of prediction is not provided when (a) insufficient baseline data are collected thus preventing a clear understanding of behavior prior to intervention, (b) substantial variability exist in both baseline and intervention data resulting in substantial overlap in data between conditions, (c) baseline data trending in the direction it would be expected to trend with intervention implementation, and (d) failure of behavior to change when the intervention is implemented. The second element of single subject designs is verification, which provides evidence that an intervention is the cause of changes in behavior. Verification can be demonstrated by either withdrawing the intervention and observing a return to baseline levels of behavior or evidence that behaviors not yet targeted through intervention (either engaged in by the same person or others within the study) have not changed from their baseline levels. The final element used to demonstrate the effects of intervention is replication. Replication is demonstrated either by behavior changing in the direction intended with the reinstatement of the intervention or intervention effects being replicated across other behaviors, individuals, or settings (Richards et al. 1999).

Although it is not reasonable to expect RTI teams to evaluate interventions in the same manner as tightly controlled ABA studies, there is much that can be learned from these analytic procedures. First, RTI teams would benefit from developing a greater understanding of the need to collect baseline data. Universal screening data can serve as a form of baseline data as it provides evidence that a student has failed to respond to tier 1 instruction and, thus, their level of academic performance falls below their peers. Universal screening data alone does not, however, provide information regarding the variability and trend in student behavior prior to intervention implementation. Second, RTI teams would benefit from addressing the question of whether an intervention is effective instead of only addressing the question of whether a student is making adequate progress.

Answering the question of whether intervention is effective for a student would require teams to measure behaviors that are directly targeted by the selected intervention, thus, allowing RTI teams to evaluate intervention effects within shorter periods of time. It would also allow RTI teams to use the data they collect for making informative decision regarding whether a student has mastered the skill(s) being targeted by the intervention and, thus, when the target of intervention needs to be modified. Ultimately, answering the question regarding whether intervention is improving student behavior by collecting data on the skills being targeted through intervention will result in (a) implementation of ineffective interventions for shorter periods of time, (b) increases in the rate of gains made by students as RTI teams can be more responsive to the changing instructional needs of students, and (c) RTI teams being able to better predict students' responsiveness to instruction as they observe the rate at which students acquire the skills necessary to make improvements on generalization measures.

Conclusion

Assessment within an RTI framework is a relatively new endeavor for many schools. Unlike the IQ-achievement discrepancy model, much of the decision-making is more complex. The decision of whether a student should qualify for special education services is not based on a simple discrepancy between two constructs. Rather, it is based upon analyses of data and whether those data suggest the student is responding adequately to intervention. Implementation of RTI models are also complicated by the fact that poor classroom behavior and academic performance may be due to the instructional environment and decisions made by the RTI team. A student's response to instruction is dependent upon the quality of tier 1 instruction, accurate identification and measurement of targeted behaviors, the selection of an appropriate evidenced-based intervention, and implementation of the intervention with fidelity.

Although the empirical literature on RTI is still emerging, schools can reference the extensive research available on the principles of ABA, which serve as a foundation for RTI. The science of ABA, with its focus on measuring behavior and the impact of the environment on behavior, provides a scientific foundation from which schools can draw upon to ensure that their instructional environment will maximize achievement for students of all abilities. In 1986 when describing differences between traditional assessment and ABA, Hayes, Nelson, and Jarrett wrote, "Rather than seek the pure group first, perhaps we should let treatment responsiveness or other functional effects select our groups for us" (p. 499). It is this principle upon which RTI frameworks are built. We must not place students into categories based upon measures of mental constructs, but rather we must provide instruction to students based upon data that directly informs us of their instructional needs.

References

Achenbach, T. (1991). *Manual for the child behavior checklist/4- 18 and 1991 profile*. Burlington: University of Vermont Department of Psychiatry.

Alter, P., Brown, E., & Pyle, J. (2011). A strategy-based intervention to improve math word problem solving skills of students with emotional and behavioral disorders. *Education & Treatment of Children, 34*(4), 535–550. doi:10.1353/etc.2011.0028.

Ardoin, S. P. (2006). The response in response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43*(6), 713–725.

Ardoin, S. P., & Daly, E. J., III (2007). Introduction to the special series: Close encounters of the instructional kind-how the instructional hierarchy is shaping instructional research 30 years later. *Journal of Behavioral Education, 16,* 1–6.

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16,* 55–70.

Ardoin, S. P., Roof, C. M., Klubnik, C., & Carfolite, J. (2008). Evaluating curriculum-based measurement from a behavioral assessment perspective. *The Behavior Analyst Today, 9,* 36–48.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18. doi:10.1016/j.jsp.2012.09.004.

Belfiore, P. J., Skinner, C. H., & Ferkis, M. A. (1995). Effects of response and trial repetition on sight-word training for students with learning disabilities. *Journal of Applied Behavior Analysis, 28,* 347–348.

Bloom, S. E., Iwata, B. A., Fritz, J. N., Roscoe, E. M., & Carreau, A. B. (2011). Classroom application of a trial- based functional analysis. *Journal of Applied Behavior Analysis, 44*(1), 19–31.

Brown-Chidsey, R., & Steege, M. W. (2005). *Response to intervention*: *Principles and strategies for effective practice*. New York: Guilford.

Call, N. A., Pabico, R. S., & Lomas, J. E. (2009). Use of latency to problem behavior to evaluate demands for inclusion in functional analyses. *Journal of Applied Behavior Analysis, 42,* 723–728.

Carr, E. G., & Durand, M. V. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*(2), 111–126.

Carr, J. E., & Sidener, T. M. (2002). On the relation between applied behavior analysis and positive behavioral support. *The Behavior Analyst, 25*(2), 245.

Cates, G. L., & Rhymer, K. N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education, 12,* 23–34.

Cooper, J. O. (1982). Applied behavior analysis in education. *Theory into Practice, 21*(2), 114–118.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). New Jersey: Pearson.

Daly, E. J., III, Martens, B. K., Witt, J. C., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26,* 554–574.

Daly, E. J., III, Murdoch, A., Lillenstein, L., Webber, L., & Lentz, F. E. (2002). An examination of methods for testing treatments: Conducting brief experimental analyses of the effects of instructional components on oral reading fluency. *Education & Treatment of Children, 25,* 288–316.

Daly, E. J., III, Chafouleas, S. M., & Skinner, C. H. (2005). *Interventions for reading problems*: *Designing and evaluating effective strategies*. New York: Guilford.

Daly, E. J., III, Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36*(4), 562–581.

Drasgow, E., & Yell, M. L. (2001). Functional behavioral assessments: Legal requirements and challenges. *School Psychology Review, 30,* 239–251.

Eckert, T. L., Ardoin, S. P., Daly, E. J., III, & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35,* 271–281.

Fisher, W. W., Groff, R. A., & Roane, H. S. (2011). Applied behavior analysis: History, philosophy, principles, and basic methods. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 3–33). New York: Guilford.

Fletcher, J., Coulter, W. A., Reschly, D., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54*(2), 304–331.

Fletcher, J. M., Denton, C., & Francis, D. J. (2005). Validity of alternative approaches for the identification of learning disabilities: Operationalizing unexpected underachievement. *Journal of Learning Disabilities, 38,* 545–552.

Galassi, J. P., & Perot, A. R. (1992). What you should know about behavioral assessment. *Journal of Counseling and Development, 70*(5), 624–631.

Geiger, K. B., Carr, J. E., & LeBlanc, L. A. (2010). Function-based treatments for escape-maintained problem behavior: A treatment-selection model for practicing behavior analysts. *Behavior Analysis in Practice, 3,* 22–32.

Goldstein, A. P., & Martens, B. K. (2000). *Lasting change*: *Methods for enhancing generalization of gain*. Champaign: Research Press.

Gresham, F. M. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review, 33*(3), 326–343.

Haring, N. G., & Eaton, M. D. (1978). Systematic procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R*: *Research in the classroom* (pp. 23–40). Columbus: Charles E. Merril.

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1986). *Evaluating the quality of behavioral assessment. Conceptual foundations of behavioral assessment* (pp. 463–503). New York: Guilford.

Hine, J. F., & Ardoin, S. P. (2011, May). *Using discrete trials to increase the feasibility of assessing student problem behavior*. Paper presented at the 37th annual conference of the Association of Applied Behavior Analysis International. Denver, CO.

Hosp, J. L., & Ardoin, S. P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention, 33*(2), 69–77.

Hudson, T., Hinkson-Lee, K., & Collins, B. (2013). Teaching paragraph composition to students with emotional/behavioral disorders using the simultaneous prompting procedure. *Journal of Behavioral Education, 22*(2), 139–156. doi:10.1007/s10864-012-9167-8.

Individuals with Disability Education Act Amendments of 1997 [IDEA]. (1997). http://thomas.loc.gov/home/thomas.php. Accessed 10 Dec 2013.

Individuals with Disabilities Education Improvement Act of 2004. (2004). Pub. L. No. 108–446.

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209.

Ledford, J. R., Gast, D. L., Luscre, D., & Ayres, K. M. (2008). Observational and incidental learning by children with autism during small group instruction. *Journal of Autism and Developmental Disorders, 38*(1), 86–103.

LeGray, M. W., Dufrene, B. A., Sterling-Turner, H., Joe Olmi, D. D., & Bellone, K. (2010). A comparison of function-based differential reinforcement interventions for children engaging in disruptive classroom behavior. *Journal of Behavioral Education, 19*(3), 185–204. doi:10.1007/s10864-010-9109-2.

Marston, D. (2005). Tiers of intervention in responsiveness to intervention: Prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities, 38,* 539–544.

Martens, B. K., & Ardoin, S. P. (2010). Assessing disruptive behavior within a problem-solving model. In G. G. Peacock, R. A. Ervin, E. J. Daly, & K. W. Merrell (Eds.), *Practical handbook in school psychology*: *Effective practices for the 21st century* (pp. 157–174). New York: Guilford.

Martens, B. K., Eckert, T. L., Bradley, T. A., & Ardoin, S. P. (1999). Identifying effective treatments from a brief experimental analysis: Using single case design elements to aid decision-making. *School Psychology Quarterly, 14,* 163–181.

McDougal, J. L., Graney, S. B., Wright, J. A., & Ardoin, S. P. (2010). *RTI in practice*: *A practical guide to implementing effective evidence-based interventions in your school*. Hoboken: Wiley.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to Intervention screening and progress monitoring practices in 41 local school. *Learning Disabilities Research & Practice, 24,* 186–195.

Mesmer, E. M., & Duhon, G. J. (2011). Response to intervention: Promoting and evaluating generalization. *Journal of Evidence-Based Practices for Schools, 12*(1), 75–104.

Miller, F. G., & Lee, D. L. (2013). Do functional behavioral assessments improve intervention effectiveness for students diagnosed with ADHD? A single-subject meta-analysis. *Journal of Behavioral Education, 22*(3), 253–282. doi:10.1007/s10864-013-9174-4.

Nelson, R. O., & Hayes, S. C. (1979a). Some current dimensions of behavioral assessment. *Behavioral Assessment, 1,* 1–16.

Nelson, R. O., & Hayes, S. C. (1979b). The nature of behavioral assessment: A commentary. *Journal of Applied Behavior Analysis, 12*(4), 491–500.

Noell, G. H., Call, N. A., & Ardoin, S. P. (2011). Building complex repertoires from discrete behaviors by establishing stimulus control, behavioral chains, and strategic behavior. In W. Fisher, C. Piazza, & H. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 250–269). New York: Guilford.

Parker, D., & Kamps, D. (2011). Effects of task analysis and self-monitoring for children with autism in multiple social settings. *Focus on Autism and Other Developmental Disabilities, 26*(3), 131–142.

Positive Behavioral Interventions and Supports. (2013). Technical Assistance Center on Positive Behavioral Interventions and Supports. Office of Special Education Programs, 2013. Web. <http://www.pbis.org>. Accessed 10 Dec 2013.

President's Commission on Excellence in Special Education. (2002). *A new era*: *Revitalizing special education for children and their families*. Jessup: ED Pubs.

Reynolds, C. R., & Kamphaus, R. W. (2006). *BASC-2: Behavior assessment system for children* (2nd ed.). Upper Saddle River: Pearson Education, Inc.

Richards, S., B., Taylor, R. L., Ramasamy, R., & Richards, R. Y. (1999). *Single subject research*: *Applications in educational and clinical settings*. San Diego: Singular Publishing Group, Inc.

Roane, H. S., Rihgdahl, J. E., Kelley, M. E., & Glover, A. C. (2011). Single-case experimental design. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 132–150). New York: Guilford.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36,* 601–617.

Shumate, E. D., & Wills, H. P. (2010). Classroom-based functional analysis and intervention for disruptive and off-task behaviors. *Education and Treatment of Children, 33,* 23–48.

Solnick, M., & Ardoin, S. P. (2010). A quantitative review of functional analysis procedures in public school settings. *Education and Treatment of Children, 33,* 153–175.

Sparks, R. L., & Lovett, B. J. (2009). Objective criteria for classification of postsecondary students as learning disabled. *Journal of Learning Disabilities, 42*(3), 230–239.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

Sugai, G., & Horner, R. H. (2002). Introduction to the special series on positive behavior support in schools. *Journal of Emotional & Behavioral Disorders, 10*(3), 130.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R., III, Wickham, D., Reuf, M., & Wilcox, B. (2000). Applying positive behavioral support and functional behavioral assessment in schools. *Journal of Positive Behavioral Interventions, 2,* 131–143.

Szadokierski, I., & Burns, M. K. (2008). Analogue evaluation of the effects of opportunities to respond and ratios of known items within drill rehearsal of Esperanto words. *Journal of School Psychology, 46*(5), 593–609.

Tkachuk, G., Leslie-Toogood, A., & Martin, G. L. (2003). Behavioral assessment in sport psychology. *The Sport Psychologist, 17,* 104–117.

Vollmer, T. R., & Iwata, B. A. (1992). Differential reinforcement as treatment for behavior disorders: Procedural and functional variations. *Research in Developmental Disabilities, 13*(4), 393–417.

Learning Disabilities/Special Education

John L. Hosp, Sally Huddle, Jeremy W. Ford and Kiersten Hensley

Special education has a long tradition of valuing a continuum of placements and services to meet individual students' needs (Deno 1970). The alignment with response to intervention (RTI) and the provision of evidence-based intervention is strong. In this chapter, we discuss the research base and characteristics of RTI as they relate to students with learning disabilities (LD) and special education services in general. We discuss RTI as both a system of educational service delivery as well as a method for determining eligibility for special education services as a student with LD. Both definitions and approaches are important to examining the intersection of RTI and special education and there is a good amount of overlap.

The authors begin this chapter with definitions of some of the key terms that are important to the treatment of the topic throughout the rest of the chapter. Next, the research base in relation to some key components of RTI and how they link to the needs of students with LD and the provision of special education services are presented. These include (a) eligibility and exit determinations, (b) tiered delivery of services, (c) evidence-based practices (EBPs), (d) universal screening, and (e) progress monitoring. Finally, some directions for future research in this area that are derived from prior research and gaps therein as well as implications of the current research base for the implementation of RTI are provided.

J. L. Hosp (✉)
University of Iowa, Iowa, USA
e-mail: john-hosp@uiowa.edu

S. Huddle
Iowa Wesleyan College, Iowa, USA

J. W. Ford
Boise State University, Idaho, USA

K. Hensley
Minnesota State University, Mankato, Minnesota, USA

Definitions

- *Aptitude–achievement discrepancy*: A method for identifying students with LD. It includes comparing a student's performance on a standardized, norm-referenced achievement test to a test of cognitive ability. A student is determined to have a LD if his/her achievement score is less than would be predicted given his/her cognitive ability. The RTI framework can be used as an alternative method for identifying students with LD.
- *Child Find*: A federal mandate that requires school districts to locate, identify, and evaluate children with disabilities, from birth to age 21, to ensure provision of a free and appropriate public education through special education and/or related services.
- *Curriculum-based measurement (CBM)*: A measurement system which uses instruments and metrics highly associated with the curriculum a student is expected to learn. Thus, little inference is required regarding the determination of students' skill level.
- *Evidence-based practices (EBPs)*: Instructional practices that are supported by high-

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_4

quality research to achieve their desired effect (typically improving student outcomes).

- *Fidelity of implementation*: The extent to which an intervention is implemented as intended.
- *Individuals with Disabilities Education Act (IDEA)*: A federal special education law that requires public schools to serve students with disabilities as determined by their individualized education program (IEP).
- *Progress monitoring*: Repeated assessment of students' academic skills and/or behavior to assist in determining whether they are making progress toward meeting expectations. Measures for progress monitoring are often the same as for universal screening. The frequency with which they are administered is dependent upon the resources allocated to the student.
- *Slope*: When students receive supplemental or intensive resources, their progress is monitored frequently and graphed. Slope refers to the line of best fit when considering at least 8–15 data points and represents the student's progress toward meeting expectations.
- *Tier/level*: Distinguishable levels of service or resources within a multi-tiered system of supports. The levels can be distinguished by characteristics such as the type of instruction provided, the size of grouping for instruction, and the amount of time service is provided.
- *Tier 1/core*: Instruction and educational resources (including materials and personnel) provided to all students within a school.
- *Tier 2/supplemental*: Instruction and educational resources (including materials and personnel) provided to students who are at risk for school failure. Supplemental is provided in addition to, rather than in lieu of, core.
- *Tier 3/intensive*: Interventions and educational resources (including materials and personnel) provided to students who did not show sufficient progress in tiers 1 and 2. In some RTI models this is, in whole or part, special education services. It should also be provided in addition to, rather than in lieu of, core. In some RTI models, it supplants supplemental services.
- *Universal screening*: Assessment of all students' academic skills and/or behavior, typically three times per year, in order to identify students at-risk for not meeting expectations at the end of the year.

Research Linking LD/Special Education to RTI

Eligibility and Exit

Students must go through a multi-gated process in order to receive special education services. This process is intended to ensure that only students who have a disability *and* a need for specialized instruction are identified for services. The first step in the process of determining eligibility for special education services is identifying a child as one who needs specialized services. This occurs in one of two ways: Child Find or referral. Child Find is a process required by IDEA for schools to find children aged 3–21 suspected of having a disability.[1] Referral often comes from a teacher or other educational service provider, but can come from a parent requesting that his/her child be evaluated. The next step involves a full and individual evaluation to examine all educational areas for the child. The last step is to decide whether the results from evaluation determine eligibility. The steps are the same regardless of age or location, but the way education professionals go about evaluating and determining eligibility can differ.

The operationalization of the eligibility determination process for special education services has been debated since the inception of the Education for All Handicapped Children Act (EHA) in 1975. Federal definitions of disabilities did not include guidelines on how to identify those with disabilities, particularly those with LD.[2] Seeing

[1] In some states, birth-3 Child Find is conducted by the Department of Health and Human Services (or equivalent) rather than the Department of Education.

[2] In the federal legislation and regulations, LD is referred to as specific learning disabilities (SLD); however, for consistency we use the more generic term of learning disabilities (LD) throughout this chapter.

how the absence of guidelines might be problematic, the US Office of Education (1976) outlined procedures for the identification of LD. These procedures included a formalized discrepancy model:

> A specific learning disability may be found if a child has a severe discrepancy between achievement and intellectual ability in one or more of several areas: oral expression, written expression, listening comprehension or reading comprehension, basic reading skills, mathematics calculation, mathematics reasoning or spelling. A "severe discrepancy" is defined to exist when achievement in one or more of the areas falls at or below 50 % of the child's expected achievement level, when age and previous educational experiences are taken into consideration (p. 52405).

This specified formula for determining severe discrepancy was met with so much negative feedback that it was dropped from the regulation in U. S. Department of Education (1977), but the concept of discrepancy remained, leaving states and districts to determine their own procedures for using discrepancy to identify LD and determine eligibility for special education services.

One widely used model is the IQ–achievement discrepancy model, which examines the difference between estimated and actual performance of a student in areas related to LD, with the estimated performance typically based on IQ (Kavale 2002). This approach has been widely disputed, cited as being a root cause of inconsistent identification. Varying ways to calculate IQ discrepancy, differences in determining the magnitude of the discrepancy, and a variety of IQ and achievement measures used have led to inconsistencies in identification of LD between and within states (Fuchs and Fuchs 2006; Reschly and Hosp 2004).

The discrepancy model has also led to difficulties in differentiating LD from low achievement (LA). Ysseldyke et al. (1982) studied the difference between students identified with LD and low achievers, finding that there was little to no difference in scores on a variety of psychometric measures. Overlap in performance between the groups ranged from 82 to 100 %, illustrating that the discrepancy model was not effective in the appropriate identification of LD.

Another criticism of using a discrepancy model is that it often leads to struggling learners having to "wait to fail" in order to become discrepant enough to receive help (Gresham 2002). The selection of students who are screened often falls to the teacher, who often screens through the use of observation, instead of through the use of a validated measurement tool. This can lead to misidentification, unnecessary failure, and loss of instructional time (Vaughn and Fuchs 2003).

It is this concern that led to what some consider as the regulatory foundation of RTI (Kovaleski and Prasse 1999; Hosp 2001). In the 1997 reauthorization of IDEA, the exclusionary criteria for special education eligibility included a provision that the main determinant of the student's poor academic performance could not be lack of instruction in reading or mathematics. At that time, little guidance was given about how to measure this lack of instruction (Hosp 2001) and no mention was included of EBP.

With the reauthorization of IDEA in 2004 came additional guidance on identifying students with LD. The difference came in the requirement that states allow the use of the RTI process and cease the requirement of IQ discrepancy when identifying students with LD. There are variations of how RTI is included and operationalized in LD eligibility among states, but most rely on some variant of the dual discrepancy model (Fuchs and Fuchs 1997; 1998). This model uses universal screening and/or progress monitoring data to compare a student's level of performance and rate of progress to criteria for acceptable standards. These criteria can be either norm referenced or empirically derived to predict performance on an outcome instrument such as the state-mandated accountability measure. As such, student level of performance can be either adequate or discrepant as can the student's rate of progress. Figure 1 illustrates the four decisions that can be derived from comparing both level and rate to criteria ordered from the least to most discrepant warranting the most intensive interventions.

Using RTI to determine eligibility removes or reduces many of the issues of using IQ discrepancy. The RTI model has been shown to reduce the

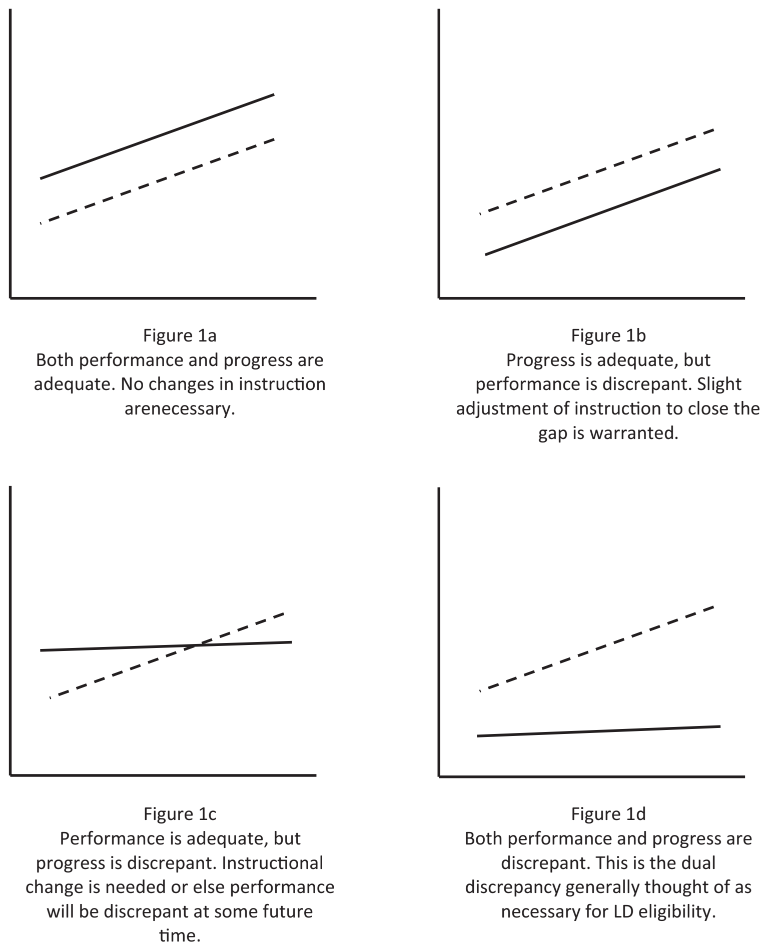


Fig. 1 The four decisions within a dual-discrepancy framework. The dotted line represents the standard; the solid line student performance. **a** Both performance and progress are adequate. No changes in instruction are necessary. **b** Progress is adequate, but performance is discrepant. Slight adjustment of instruction to close the gap is warranted. **c** Performance is adequate, but progress is discrepant. Instructional change is needed or else performance will be discrepant at some future time. **d** Both performance and progress are discrepant. This is the dual discrepancy generally thought of as necessary for LD eligibility. *LD* learning disabilities

number of students referred for special education in the early elementary grades. Ikeda and Gustafson (2002) found a 25 % reduction in referrals. Tucker and Sornson (2007) found that the implementation of RTI through instructional support teams was especially effective in reducing special education placement for minority students. The use of an RTI model also reduces the bias that exists in the referral process. Variability in referral often comes from a classroom teacher's view of how students should perform and respond to instruction, but the assessment process can serve to reduce disproportionality in referrals (Hosp and Reschly 2003). This also holds true for the systematic procedures in universal screening (Vaughn and Fuchs 2003).

The confounding of LD and LA can also be lessened through the RTI process and the ability to evaluate skill versus performance deficits. The process of evidence-based instruction, universal screening, and progress monitoring is built to identify those students who do not respond sufficiently to intervention. These are often considered the students with LD. The students who are LAs will respond to intervention and not be referred for special education services (Kavale et al. 2005).

RTI also addresses the issues of "wait to fail" often seen within discrepancy models. Students participate in evidence-based instruction and universal screening during the RTI process, leading to identification based on risk instead of waiting for a certain amount of discrepancy. This allows for earlier identification of students in need, along with the provision of evidence-based intervention and progress monitoring for all students at tier 2 or 3, even before being identified with LD (Vaughn and Fuchs 2003).

Going back to the required process of determining student eligibility for special education services, the RTI process fits within all the steps. First, referral must occur. This happens naturally through the screening process. Second, a full and individual evaluation must take place. Data collected from interventions and progress monitoring can be used to complete the evaluation, but additional information may need to be collected. Using all of the data and information collected during the process leads to the ability to make decisions regarding eligibility.

Researchers have completed numerous studies on how to make decisions on eligibility, but very little has been done on exiting students from special education services. It is too easy for a student to get stuck in the special education system, which has been designed to recognize discrepancy, but not necessarily to recognize when discrepancy has been diminished (Powell-Smith and Stewart 1998). Planning for exit should be part of the discussion at all times. RTI models facilitate the review of student data in comparison to expected general education performance through the continuous collection of progress monitoring data, leading to the return to general education when performance is comparable to their peers in general education (Grimes 2002). Ideally, this would be an indicator of the health of the special education system.

Tiers of Service Delivery

Fletcher and Vaughn (2009) describe the goals of RTI to improve academic and behavioral outcomes for all students and provide data to determine eligibility for special education services using a student's response to instruction instead of the discrepancy between their cognitive abilities and achievement. The reauthorization of IDEA in 2004 provided schools with the opportunity to use a student's response to evidence-based instruction as a method for identifying students eligible for and in need of special education services. This model required schools to have distinguishable levels of instruction and intervention and opened the door for large-scale adoption of RTI as a way for schools to provide multi-tiered systems of support for all students.

RTI's system of support includes universal screening, progress monitoring, and evidence-based instruction for all students, but RTI is largely defined by its multiple tiers of support. Even though there are different models with a varying number of tiers of support, it is widely accepted that there must be at least three tiers (Bradley et al. 2007). The tiers are aimed at both prevention of academic and behavioral difficulties and early identification of students at risk for such difficulties. Table 1 describes the type of instruction and assessments used at each level.

Table 1 Characteristics of tiers of RTI

Level	Description of instruction	Description of assessment
Tier 1/core	Evidence-based instruction delivered with fidelity in the general education setting to all students	Universal screening to identify students at risk for school failure
Tier 2/supplemental	Small group evidence-based instruction provided for 3–5 days per week for 8–10 weeks to students determined to be at risk for school failure. It is in addition to tier 1	Progress monitoring (weekly or biweekly) to determine students' response to the intervention provided
Tier 3/intensive	Individualized interventions including, but not limited to, special education services. Provided 1:1 or in small groups (≈3), generally 5 days per week for 45–90 min each day	Progress monitoring (at least weekly) to determine students' response to the intervention provided. Diagnostic assessment about specific student needs to adjust intervention accordingly

RTI response to intervention

With the widespread adoption of RTI came the implementation of some practices that had not been empirically validated. These practices included one-stage screenings to determine academic risk and the prerequisite that all students must show inadequate response to tier 1 *and* 2 instruction before being allowed to participate in tier 3 intensive interventions or special education services (Fuchs et al. 2012). Recent research on RTI's multi-tiered system of support indicates multi-stage screenings that may help educators reduce unnecessary tier 2 interventions for some students and provide other students intensive tier 3 intervention without waiting for them to fail in tier 2 (Fuchs et al. 2012). Multi-stage screening makes more efficient use of school resources and maximizes students' opportunities for success by using a type of direct route screening system (Jenkins et al. 2007). Vaughn et al. (2010) determined that it may be inappropriate to place some students in lower intensity tier 2 interventions when they are significantly below grade-level expectations and unlikely to close the gap without high-intensity tier 3 interventions.

There has been some controversy over the role of special education within the tiers of RTI. Included in this controversy is if special education should be part of tier 3 or operate outside the tiers and if the purpose of RTI is to prevent special education placement or prevent school failure (Fuchs et al. 2012). Recent research has demonstrated that even with adequate tier 1 and tier 2 intervention and instruction, approximately 5% of students will continue to be in need of more intensive and specialized tier 3 interventions, and these interventions are best delivered by educators with specialized expertise such as special education teachers (Fuchs et al. 2012). Tier 3 evidence-based instruction for students with LD must be the most intense in terms of explicit instruction, opportunities for students to respond and engage, time in intervention, and expertise of the instructor (Vaughn et al. 2010). This requires a detailed understanding of what EBPs are.

Evidence-Based Instruction

In order for an instructional practice or program to be called an EBP, it must meet prescribed standards in several areas such as research design, impact on students, and the source of evidence (Cook and Odom 2013). Gersten et al. (2005) and Horner et al. (2005) provided standards for identifying EBP's within special education and these standards are used on an ongoing basis to identify EBP's for students receiving special education services within an RTI model (Cook and Odom 2013).[3]

EBPs are a hallmark of the RTI model. RTI specifically relies on a student's response to EBPs. However, it is important to note that EBPs are not guaranteed to work with every student. Practices identified as EBPs are likely to show improved outcomes for the majority of students but not all students will respond equally to every EBP. Researchers estimate approximately 5% of students will not respond sufficiently to well implemented EBPs at tiers 1 and 2 and need more intensive tier 3 services (Fuchs et al. 2012; Torgeson 2000). Cook and Odom (2013) suggest that because not all educational practices have been thoroughly researched, educators may need to use practices that have not been deemed EBPs. However, this should be done with caution and only when students have shown poor response to previous implementation of EBP's. This practice would most likely occur during tier 3 special education services, and educators should not use practices that have been shown by research to be ineffective.

Once an instructional practice or program has sufficient evidence to be labeled an EBP, there remains the important step of implementation. In order to have improved outcomes for students with disabilities, educators must not only

[3] Although it is important for educators to understand quality indicators of research to indicate evidence-based practices, there are several organizations that systematically apply these and other criteria to provide recommendations for which are evidence based. These include the What Works Clearinghouse, the National Center on Intensive Interventions, the Center on Instruction, the Best Evidence Encyclopedia, and others.

Table 2 Evidence-based practices for students receiving special education services

–	Academic	Behavior
Source	Swanson (2000; 2001)	Landrum et al. (2003) Sutherland et al. (2008)
Evidence-based practices	1. Daily review 2. Statement of objective 3. Teacher presents new material 4. Guided practice 5. Independent practice 6. Active presentation 7. Explicit strategy instruction 8. Explicit practice (review, repeated practice, feedback) 9. Examples 10. Demonstrations 11. Visual prompts 12. Strategy cues	1. Reinforcement 2. Precision requests 3. Behavioral momentum 4. Time-out 5. Response cost 6. Group contingencies 7. Monitoring of student performance 8. Classroom management (including clear expectations and routines) 9. Pre-correction strategies

choose EBPs but also implement them with fidelity (Sanetti and Kratochwill 2009). The relation between efficacy and implementation is what allows educators to determine individual students' RTI. In order to enhance both the efficacy and implementation of EBP practices within tier 3 instruction, special educators should look to implement specific practices and programs. These practices and programs should have components which are easy to identify and have been validated by research but at the same time are flexible enough to allow for adaptation based on students' need and can be used with different types of students (Harn et al. 2013). There are several EBPs meeting these criteria that are detailed in Table 2.

Within tier 3 special education services, there are also EBPs to guide intervention intensity. These include instruction provided (Vaughn et al. 2010):

- In small groups (e.g., 3–5 students),
- For extended time (e.g., 60 min per day),
- Over extended time periods (e.g., 20–30 weeks),
- That is both explicit and systematic (e.g., follows a prescribed scope and sequence of skills).

For younger students, interventions may not need to be as intense in regard to the amount of time spent in intervention. For older students, interventions may need to be more intense with more time spent in intervention because academic problems are more difficult to remediate when the student is older (Juel 1988).

Universal Screening

Comprehensive school assessment systems include the use of universal screening (Salvia et al. 2012). Universal screening seeks to answer the question: How well are students (i.e., at the specific class, grade, school, or district level) meeting academic and/or behavioral expectations? Data for universal screening can be collected individually or in groups and commonly occurs three times during the school year (e.g., fall, winter, and spring). Appropriate measures for universal screening are quickly administered, quickly scored, and able to provide information for valid inferences regarding what is being measured (Hosp and Ardoin 2008). Universal screening is also a key component in problem-solving models such as the RTI framework (Batsche et al. 2005).

In discussing universal screening within the context of resource allocation, Ikeda et al. (2008) identify two primary purposes, helping to determine (a) whether or not a school's core instruction is meeting the vast majority of students' needs and (b) which students may require additional (i.e., supplemental or intensive) resources. The former purpose relates to universal screening being beneficial to all students while the latter highlights how the practice helps meet the needs of students at risk for not meeting desired outcomes. Students with LD are among those most likely to be at risk for not meeting expectations without additional support.

Given the purposes of universal screening, it is vital to note that the practice is not intended to identify specific explanations for why students may be having difficulty. In addition, it should also be clear that universal screening does not provide data for designing specific instructional interventions. Instead, an accurate representation of universal screening is that it provides a snapshot of student performance given the current state of how resources are allocated. As mentioned previously, universal screening practices are important for all students as well as students with LD.

Two features of universal screening highlight its utility for use with all students. The first feature is how resulting data can be used to show a student's relative standing regarding his/her likelihood for meeting desired expectations compared to peers. In order to make such a comparison, it is necessary that all students participate in universal screening. The second feature that highlights universal screening being a process for use with all students is how resulting data can be used to observe student performance in relation to a predetermined benchmark. This latter feature is an example of using a criterion-referenced approach.

A criterion-referenced approach to universal screening can be used to indicate the overall "health" of a school. In such an approach, a healthy school would be one with 80% of students being observed to be likely to meet desired academic/behavioral outcomes with the allocation of only core resources. In addition, a healthy school would have many students receiving supplemental and intensive resources also being observed to be likely to meet desired academic/behavioral outcomes. Having many students receiving more than core resources observed at or above a predetermined benchmark is due, in part, to efficient and effective pairing of appropriate interventions with the students who would most benefit (Hosp and Ardoin 2008).

A critical element in determining the quality of a measure for universal screening is its predictive validity (Hosp et al. 2011). While teachers do a good job of identifying students likely to have problems meeting expectations, they are also likely to fail to identify students with an equal unlikeliness (VanDerHeyden and Witt 2005). Therefore, measures with the ability to correctly distinguish true positives from true negatives are needed. True positives are cases in which students are accurately predicted to be proficient on an outcome measure by a universal screening measure. True negatives are cases in which students are accurately predicted to be nonproficient on an outcome measure by a universal screening measure. A measure's ability to accurately distinguish true positives from true negatives is commonly referred to as sensitivity and specificity respectively (Compton et al. 2010). Figure 2 illustrates these decisions and their relation to sensitivity and specificity.

Not all students identified as being unlikely to meet expectations through a universal screen-

		Outcome assessment	
		+	-
Screening Assessment	+	True Positives	False Positives
	-	False Negatives	True Negatives

Fig. 2 Classification accuracy of decisions in universal screening. True positives and negatives are those students predicted to be proficient or nonproficient who perform as expected. False positives and negatives are those students predicted to be proficient or nonproficient who do *not* perform as expected. Sensitivity is the ratio of true positive decisions to total positive performance on the outcome assessment (i.e., true positives + false negatives). Specificity is the ratio of true negatives to total negative performance on the outcome assessment (i.e., true negatives + false positives)

ing process will be students with LD. However, students with disabilities are a subgroup whose performance has historically been observed to be significantly discrepant from peers. Given the effects of early intervention to improve educational outcomes (Denton and Mathes 2003; O'Connor et al. 2005; Simmons et al. 2008) and the need for students to be identified early in order to maximally benefit from intervention (Cavanaugh et al. 2004; Vellutino et al. 2006), it is clear that universal screening practices are essential for identifying students with LD early.

Universal screening represents an effective means to assist in identifying students with LD. The practice has also been influenced by federal education policy. That is, for several years, federal educational policy has focused on increasing the skills of students with LD. This has been done by including students with disabilities in taking tests of achievement as well as the development of accountability systems which include students with disabilities. Proponents of such accountability systems have viewed them as "the critical lever for enhancing opportunities to learn and thus as a way of increasing equity in the educational system" (Resnick et al. 2003, p. 4). Further, it has been suggested that reducing discrepancies in subgroup achievement, as well as increasing overall student achievement, is an approach for pursuing social justice and economic well-being at the community, state, and national levels (Roach and Frank 2007). The influence of federal educational policy in regard to students with disabilities is an important one.

In reaction to the criticism of US schools in *A Nation at Risk* (Gardner 1983) the educational standards movement began. At first, students with disabilities were typically excluded from measuring performance toward standards. This began to change late in the twentieth century as federal education policy began to include students with disabilities in assessment of academic standards.

Beginning with the 1997 reauthorization of IDEA, students with disabilities were required to be included in assessments of standards as well as have access to the same curriculum as their peers without disabilities. The 2002 reauthorization of the Elementary and Secondary Education Act (ESEA) required all students, including most with disabilities, to be included in the assessment of challenging standards established by states. Students with significant cognitive disabilities were allowed to participate on an alternate assessment, but one in which the academic standards of the highest caliber possible were reflected. When IDEA was reauthorized in 2004 as the Individuals with Disabilities Education Improvement Act (IDEIA), it strengthened the concepts of alternate assessment and general education curriculum access. It also incorporated the academic accountability ideas of the 2002 reauthorization of ESEA.

Success in implementing RTI necessitates accurate identification of students who are at-risk for not meeting desired outcomes (e.g., Compton et al. 2006; Fuchs and Fuchs 2007; McCardle et al. 2001; VanDerHeyden et al. 2007). Universal screening represents the primary way for identifying such students (Glover and Albers 2007). Universal screening results in data and, within the RTI framework, data-based decision making is essential (Ball and Christ 2012).

Data obtained from the use of curriculum-based measurement (CBM; Hosp et al. 2007) is one method for effective universal screening practices. Specifically, measurement of student level of performance (i.e., performance at a static point in time) is one of the two critical features of the RTI framework (Fuchs 2004). Using CBM to measure skill level at a specific time (i.e., fall, winter, and spring) is an example of universal screening using CBM in the RTI framework. The second critical feature identified by Fuchs (2004) in the RTI framework is measurement of the slope of student performance. This critical feature is related to the concept of progress monitoring.

Progress Monitoring

Progress monitoring includes a comprehensive system of brief, repeated assessments which are used to help determine whether or not students are progressing in the curriculum as to meet desired expectations (Stecker et al. 2008). Resulting data from progress monitoring are graphed and a line of best fit is used to indicate a student's rate of improvement. In this way, progress monitoring data can be used to indicate current level of student's performance as well as growth. Many measures used for universal screening can also be used for progress monitoring (Hosp 2011), but the frequency in which progress monitoring measures are administered varies depending on student's skill level. This is impacted by the importance of progress monitoring.

Use of progress monitoring assists teachers and other educators in making decisions regarding students' response to evidence-based intervention across all tiers of the RTI framework (Stecker et al. 2008). Further support for progress monitoring can be found in its application through CBM. That is, research has found when teachers use CBM as a progress monitoring tool to make instructional decisions, the frequency and quality of changes to instruction significantly increases in responding to less than desired student performance (Fuchs et al. 1991).

For students with LD, the need to monitor their progress toward desired outcomes is important. Without a means to monitor progress, students may be provided an intervention which is not working for too long or have an intervention ended when it is increasing the targeted skill (Ardoin et al. 2013). Much research has been conducted on using reading CBM to monitor progress toward high-stakes state testing. For example, Hintze and Silberglitt (2005) conducted a study of a cohort of students from first through third grade examining the predictive and diagnostic accuracy of reading CBM to a high-stakes state test of reading at the end of third grade. Results suggested reading CBM to be strongly associated with performance on the state reading test across grade levels. In addition, reading CBM performance in the beginning of first grade accurately and efficiently predicted performance on the state reading test. Espin et al. (2010) also found using reading CBM to monitor student progress to be effective with secondary students. Espin et al. (2010) examined the relation between reading CBM and maze selection to a state reading test for students in eighth grade. Results suggested both measures to be consistent and accurate for predicting performance on the state reading test. In addition, while conducting a smaller exploratory follow-up study, Espin et al. (2010) found maze selection to reflect changes in growth whereas reading CBM did not. The finding of maze selection reflecting change and reading CBM failing to do so was observed for both low- and high-performing students. However, as previously stated, progress monitoring is a process that is important for all students.

When the progress monitoring process is seen as asking the question, "How are students progressing in the curriculum?" It is easy to see how progress monitoring is important for all students. That is all students need to have this question answered for them; however, the frequency in which it is answered depends on the student's performance in regard to the skill in question. For students who are progressing as expected with only the provision of core resources, it may be appropriate to only monitor their progress at the time universal screening data are collected.

It is important to note that some students identified as unlikely to meet academic/behavioral expectations as a result of universal screening may not be in need of the allocation of resources in addition to the core. A helpful feature of progress monitoring all students is found in the likeliness that universal screening may identify some students inaccurately (i.e., as a false negative as defined earlier). However, if progress monitoring data are collected for at least an additional 5 weeks, with only core instruction provided, rates of improvement may show that some students are observed to be on target to meet expected outcomes (Compton et al. 2006). Thus, important resources can be kept from being provided to students who do not require them.

Even when students do require supplemental or intensive resources to increase the likeliness of their meeting desired outcomes, many will be students without LD. Again the act of progress monitoring occurs at a frequency depending on student need. For students receiving supplemental resources, progress monitoring often occurs once a month (Hosp et al. 2007). Slope of improvement should be evaluated upon collection of 8–15 data points while students receive supplemental intervention (Stecker et al. 2008). When students receiving supplemental resources are observed to make adequate progress, a plan to release these supports to transition the student back to receiving only core resources with continued monitoring of progress should be implemented to ensure appropriate support is provided. Students who are not observed to make adequate progress should be considered for additional, more intensive allocation of resources. Although students without LD benefit from having their progress monitored, especially those receiving supplemental and intensive resources, students with LD are especially likely to see skill improvement when the practice is used (Stecker et al. 2005).

It would be an understatement to say that progress monitoring is critical for students with disabilities in the RTI framework. Given RTI's purpose of determining whether or not students are responding to evidence-based intervention, progress monitoring is key. However, the level of support (i.e., core, supplemental, or intensive) a student with a LD receives should be considered when determining how often to monitor progress. Some students with a LD may have their needs met with the provision of core resources and not require supplemental, intensive, or special education resources (Ford 2013). Thus, a student with a LD who is not receiving intensive or special education resources (perhaps through the provision of high-quality differentiated instruction and supplemental instruction) may only need their progress monitored once a month like any other student receiving supplemental resources.

While students with LD may receive core, supplemental, intensive, and/or special education resources, all students receiving the most intensive interventions will be those identified as having a LD. For students receiving more intensive resources, once or even twice a week is recommended (Hosp et al. 2007). When a student with a LD responds as desired to intensive instruction, they can be transitioned to receiving supplemental or even core resources provided their educational needs are still being met and they are making sufficient progress (D. Fuchs et al. 2008).

The role of progress monitoring has increased considerably since IDEIA 2004. This change in federal educational policy allowed for a student's response to evidence-based intervention to contribute to the identification of a LD. While this policy did not require the use of considering a student's RTI, it did make the requirement of the traditional aptitude–achievement discrepancy approach to identification unnecessary. Many consider this latter approach to be a "wait to fail model" (e.g., Gresham 2002) and the ability to use a more preventative approach such as the RTI framework was the reform focus of many schools (Tilly 2006). Thus, the need to progress monitor students at risk for not meeting desired outcomes as well as students with LD is found in federal educational policy which supports the implementation of RTI practices.

Having discussed how progress monitoring is used with all students, students with LD, and its connection with federal educational policy and subgroup accountability, it should be evident that progress monitoring and RTI are explicitly linked. The RTI framework involves a process for evaluating whether or not students are benefiting from evidence-based instruction (Stecker et al. 2008). Such instruction is framed as occurring within levels of increasingly intense levels of resource allocation. Progress monitoring (often through the use of CBM) is the means by which educators can determine if students are making sufficient progress in the curriculum (Hosp 2012). Therefore, progress monitoring is a vital component of RTI as educators determine students' instructional needs.

Areas for Future Research

Although there is a good deal of research establishing EBP in the area of LD/special education within RTI, there is still a great deal more that needs to be researched and empirically supported. The potential areas for future research, as identified by previous research as well as research literature reviews that have highlighted the gaps, are presented below organized under the categories of RTI research most relevant for LD/special education.

Eligibility and Exit

1. Much of the research on RTI for eligibility decisions exists for the lower elementary grades (K–3). Extending the implications to upper elementary (4–6) and secondary grades (7–12) is tenuous until research can be conducted demonstrating this extension is validated.
2. Although eligibility criteria have been researched, exit criteria to determine when a student no longer needs special education services and what EBP for doing this is sorely lacking. Throughout the history of special education in the USA, the focus of policy and research has almost entirely been on eligibility and entry into the system; however, to implement a system of data-based decision-making that provides a flexible continuum of services, exit (or transition to less intensive level of service) decisions must be similarly evidence based.

Tiers of Services Delivery

1. Research is needed to determine whether students with the highest levels of discrepancy between their performance and grade-level expectations benefit from earlier placement in tier 3 intensive intervention (i.e., a direct route or triage model) as compared to a systematic process through increasingly intensive levels of support (i.e., a titration model).
2. Much of the current research is focused on reading instruction; research is needed to determine how students may progress through the tiers of service delivery in other areas such as writing, mathematics, and other content areas.

Evidence-Based Practices

1. Similar to research on eligibility decisions, much of the research on EBP has been conducted at the lower elementary levels. Successful implementation of RTI in the upper grades (4–12) is dependent upon the availability of validated instruction, intervention, and measurement tools at all grade levels in all academic areas.
2. There is a continued need for high-quality research that meets the standards for EBP across all academic and behavioral domains and specifically with subpopulations of students (e.g., English language learners (ELLs)) for which there are few, if any, established EBPs.
3. There is a need for research on effective ways to provide both preservice and in-service teachers with the expertise to implement and adapt EBP to meet the needs of all students.

Universal Screening

1. Much of the research on universal screening is focused on single-stage approaches. Examination of different models of screening that include multiple stages (or gates) is important to identify the most efficient and effective.
2. In addition, universal screening measures (and approaches) still need to be improved in terms of their classification accuracy (including sensitivity and specificity) as well as their potential bias when used with different subgroups of students (e.g., ELLs).
3. The majority of research in universal screening has been conducted with reading in the elementary grades. The research base needs to be commensurate for other grade levels as well as other content areas. This includes mathematics as well as content areas such as science and social studies in the elementary grades, and specific topics such as physics and American history in secondary grades.

Progress Monitoring

1. Similar to universal screening, the research on progress monitoring has been primarily in reading for the elementary grades. Progress monitoring research needs to align with universal screening research in its focus of grade levels and content areas.
2. Also similar to universal screening, additional research is needed on potential bias across different subgroups.
3. Classification accuracy is important for progress monitoring when used as part of eligibility within a dual discrepancy framework, but there are additional measurement considerations for progress monitoring. These include the reliability and validity of slope metrics as well as the frequency of assessment and sensitivity to change.

Implications for Practice

Despite the gaps in the professional literature and the areas of additional research needed, there are still many implications of the current research base for implementing RTI and relevant for consideration of working with students with LD or who need special education services. Table 3 includes two to three implications of the current research on LD/special education for RTI implementation.

Table 3 Implications for practice for universal screening and progress monitoring practices

RTI component	Practical implications for students with LD
Eligibility and exit decisions	1. The implementation of an RTI model reduces misidentification for eligibility by systematically screening all students. 2. Educators should set criteria for exit from special education services and frequently reevaluate student performance to determine continued eligibility or the need to make an exit decision
Evidence-based practices	1. Educators should use EBP that have clearly defined components but can be flexibly adapted for use with different students. 2. Educators should carefully consider intervention intensity in order to provide students in need of tier 3 special education services with adequate amounts of intervention instruction as well as instruction that is explicit
Progress monitoring	1. Progress monitoring allows for observation of whether students with LD are making progress toward meeting expectations given current resources. 2. Progress monitoring assists in determining whether students with LD need more or less intensive resources to meet expectations. 3. Progress monitoring assists in determining if current instructional intervention is benefiting students with LD or if changes should be considered
Tiers of services delivery	1. Schools may want to spend more resources conducting multi-stage screenings for students at-risk for academic failure in order to more accurately determine which students are not at risk and which students' risk is so great they should bypass tier 2 intervention and move directly into tier 3 intensive intervention (i.e., use direct route screening). 2. Schools should consider providing higher-intensity interventions for their students with the greatest discrepancy and ensure that the teachers implementing these interventions have the necessary expertise to implement the intervention and change instruction in response to a student's needs
Universal screening	1. Universal screening is the first step in identifying students at risk for not meeting expectations, including those with LD. 2. Universal screening allows for comparison between the skill level of students with LD and their peers. 3. Universal screening allows for observation of how well students with LD are performing at the classroom, grade, and district level

LD learning disabilities, *RTI* response to intervention

Conclusions

RTI has foundations in many different disciplines and from many different sources. Federal regulations guiding assessment for the determination of eligibility of students with LD is one of these sources. Mandates for allowance of use of an RTI process in eligibility determination have expanded RTI's use across the USA. The components of RTI have direct implications for students with LD, or any students with disabilities receiving special education services. In addition to eligibility decisions, instruction and resources delivered with a tiered system, necessitate inclusion of students with LD within that system. Use of EBP as well as universal screening and progress monitoring also necessitate inclusion of students with LD. These students also require some additional considerations for assessment and intervention that might go beyond what their peers without disabilities might require.

There is an ample research base on the core components of RTI, but the directions for future research are many. When basing a system of service delivery on research, there are always new approaches to consider, new aspects of instruction or assessment to evaluate and refine, and new ways of analyzing the evidence to optimize the evidence base. This makes for an ever-changing and expanding source of implications for practice to implement the practices that are evidence based.

Special education in general is sometimes conceptualized as the final tier of an RTI system, whether that is third, fourth, or fifth. Some models consider special education as a decision and process that is implemented after provision of three levels of intervention. One thing is consistent with these models; however, special education is an important component and consideration of the provision of a continuum of services no matter how many tiers or levels there are.

References

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 19–57.

Ball, C. R., & Christ, T. J. (2012). Supporting valid decision making: Uses and misuses of assessment data within the context of RTI. *Psychology in the Schools, 49*(3), 231–244.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., & Tilly 3, W. D. (2005). *Response to intervention*. Alexandria: National Association of State Directors of Special Education, Inc.

Bradley, R., Danielson, L., & Doolittle, J. (2007). Responsiveness to intervention: 1997 to 2007. *Teaching Exceptional Children, 39*(5), 8–12.

Cavanaugh, C. L., Kim, A., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal, 2*(1), 9–21.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*(2), 327.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79,* 135–144.

Deno, E. (1970). Special education as developmental capital. *Exceptional Children, 37,* 229–237.

Denton, C. A., & Mathes, P. G. (2003). *Intervention for struggling readers: Possibilities and challenges. Preventing and remediating reading difficulties: Bringing science to scale* (pp. 229–251). Baltimore: York Press.

Espin, C., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice, 25*(2), 60–75.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30–37.

Ford, J. W. (2013). Educating students with learning disabilities in inclusive classrooms. *Electronic Journal for Inclusive Education, 3*(1).

Fuchs, L. S. (2004). Measurement research. *School Psychology Review, 33*(2), 188–192.

Fuchs, L., & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children, 30*(3), 1–16.

Fuchs, L., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13,* 204–219.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93–99.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39*(5), 14–20.

Fuchs, L. S., Hamlett, D. F. C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617–641.

Fuchs, D., Stecker, P. M., & Fuchs, L. S. (2008). Tier 3: Why special education must be the most intensive tier in a standards-driven, no child left behind world. In D. Fuchs, L. S. Fuchs, & S. Vaughn (Eds.), *Response to intervention: A framework for reading educators* (pp. 71–104). Newark: International Reading Association.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78,* 263–279.

Gardner, D. P. (1983). *A nation at risk*. Washington, D C: The National Commission on Excellence in Education, US Department of Education.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71,* 149–164.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135.

Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Response to treatment* (pp. 467–519). Mahwah: Erlbaum.

Grimes, J. (2002). Responsiveness to interventions: The next step in special education identification, service, and exiting decision making. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Response to treatment* (pp. 531–547). Mahwah: Erlbaum.

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children, 79,* 181–193.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34*(3), 372.

Horner, R. H., Carr, E. G., Halle, J., Mcgee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Hosp, J. (2001). *Assessment and determination of lack of instruction*. Washington: National Association of School Psychologists Annual Convention.

Hosp, J. (2011). Using assessment data to make decisions about teaching and learning. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 3, pp. 87–110). Washington, DC: American Psychological Association.

Hosp, J. L. (2012). Formative evaluation: Developing a framework for using assessment data to plan instruction. *Focus on Exceptional Children, 44*(9), 1.

Hosp, J. L., & Ardoin, S. P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention, 33*(2), 69–77.

Hosp, J., & Reschly, D. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education, 37,* 67–80.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.

Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review, 40*(1), 108.

Ikeda, M., & Gustafson, J. K. (2002). *Heartland AEA 11's problem solving process: Impact on issues related to special education* (Research Report No. 2002-01). Johnston: Heartland Area Education Agency 11.

Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. *Best practices in school psychology, 5,* 103–114.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review, 36,* 582–599.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Kavale, K. A. (2002). Discrepancy models in the identification of learning disability. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Response to treatment* (pp. 369–426). Mahwah: Erlbaum.

Kavale, K. A., Holdnack, J. A., & Mostert, M. P. (2005). Responsiveness to intervention and the identification of specific learning disability: A critique and alternative proposal. *Learning Disability Quarterly, 28,* 2–16.

Kovaleski, J., & Prasse, D. (1999). Assessing lack of instruction. *NASP Communiqué, 28*(4), 24–25.

Landrum, T. J., Tankersley, M., & Kauffman, J. (2003). What is special about special education for students with emotional or behavioral disorders? *Journal of Special Education, 37,* 143–156. doi:10.1177/002246 6903037003041.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research & Practice, 16*(4), 230–239.

O'Connor, R. E., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade changes in teaching and student outcomes. *Journal of Learning Disabilities, 38*(5), 440–455.

Powell-Smith, K. A., & Stewart, L. H. (1998). The use of curriculum-based measurement in the reintegration of students with mild disabilities. In M. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 254–307). New York: Guilford Press.

Reschly, D., & Hosp, J. (2004). State SLD identification policies and practices. *Learning Disabilities Quarterly, 27,* 197–213.

Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2003). Benchmarking and alignment of standards and testing. *Educational Assessment, 9,* 1–27.

Roach, A. T., & Frank, J. L. (2007). Large-scale assessment, rationality, and scientific management: The case of No Child Left Behind. *Journal of Applied School Psychology, 23,* 7–25.

Salvia, J., Ysselydke, J. E., & Bolt, S. (2012). *Assessment in special and inclusive education* (12th ed.). Boston: Cengage Learning.

Sanetti, L., & Kratochwill, T. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445–459.

Simmons, D. C., Coyne, M. D., Kwok, O. M., McDonagh, S., Harn, B. A., & Kame'enui, E. J. (2008). Indexing response to intervention a longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41*(2), 158–173.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795–819.

Stecker, P. M., Fuchs, D., & Fuchs, L. S. (2008). Progress monitoring as essential practice within response to intervention. *Rural Special Education Quarterly, 27,* 4.

Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. L. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional and behavioral disorders. *Journal of Special Education, 41,* 223–233. doi:10.1177/0022466907310372.

Swanson, H. L. (2001). Searching for the best model for instructing students with learning disabilities. *Focus on Exceptional Children, 34*(2), 1–15.

Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilities, 33,* 114–136.

Tilly, W. D. (2006). Diagnosing the Learning Enabled: The Promise of Response to Intervention. *Perspectives (Gerontological Nursing Association (Canada)), 32*(1), 29–34.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55–64.

Tucker, J. A., & Sornson, R. O. (2007). One student at a time; one teacher at a time: Reflections on the use of instructional support. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *The Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 269–278). New York: Springer.

U.S. Office of Education. (1976). *Proposed rule-making. Federal Register, 41, (230), 52404–52407*. Washington: U.S. Government Printing Office.

U.S. Office of Education. (1977). *Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. Federal Register, 42, (250), 65082–65085*. Washington, DC: U.S. Government Printing Office.

VanDerHeyden, A. M., & Witt, J. C. (2005). Quantifying context in assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34*(2), 161–183.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*(2), 225–256.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18,* 137–146.

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools, 47,* 432–444.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*(2), 157–169.

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *The Journal of Special Education, 16*, 73–85.

Prevention and Response to Intervention: Past, Present, and Future

Ann C. Schulte

The acronym "RTI" entered the educational lexicon well over a decade ago, with the initials used to represent a variety of similar terms, including "responsiveness to intervention," "response to intervention," and "response to instruction" (Fuchs et al. 2003; Lyon 1994; Speece and Case 2001; Vaughn et al. 2003). The term was first used to describe an approach to the diagnosis of reading disabilities that included brief, intensive instruction and an assessment of the student's response to that instruction (e.g., Fuchs and Fuchs 1998; Torgesen et al. 1999; Vellutino et al. 1996). RTI was intended to replace or supplement the more widely used ability–achievement discrepancy as a means of placing children in special education due to questions about the validity of the ability–achievement discrepancies as a diagnostic criterion for reading disabilities (Fletcher et al. 1994).

Although RTI's roots are in an assessment approach for the diagnosis of one disorder in one subject area (Fuchs et al. 2003), RTI has changed markedly since its first appearance in the education literature. RTI is now most usefully viewed as a service delivery approach that encompasses multiple specific models that vary on a number of attributes (Barnes and Harlacher 2008). These attributes include scope (focus on academics, behavior, or both), assessment technology (e.g., curriculum-based measurement, task analysis, norm-referenced assessment), and intervention procedures (e.g., standard protocol, problem-solving, or a combination). Although data collected as a part of RTI might provide important information for diagnosis of a reading disability, contemporary RTI is more accurately viewed as a "multilevel prevention system" that "represents a fundamental rethinking and reshaping of general education" (Fuchs et al. 2012, pp. 263–264).

Many of the key constructs in contemporary RTI can be traced directly to prevention and prevention science concepts, including early intervention to prevent later problems, universal screening, a tiered approach to delivering interventions, and a focus on evidence-based practices (Coie et al. 1993; Flay et al. 2005; Mrazek and Haggarty 1994). However, even in its earlier, more narrowly focused form as an approach to reading disability diagnosis, many elements of RTI were drawn from prevention concepts and research. For example, the problem-solving model, one of the two primary approaches to addressing children's academic difficulties in RTI (Fuchs et al. 2010), ultimately can be traced back to a behavioral consultation model (Bergan 1977) developed as part of prevention program for preschoolers targeted at preventing lifelong difficulties associated with poverty (Headstart).

The purpose of this chapter is to trace the incorporation of prevention constructs and research into RTI. The hope is that this history will enrich readers' understanding of the evolution of

A. C. Schulte (✉)
T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe, AZ 85281, USA
e-mail: ann.schulte@asu.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_5

both contemporary prevention science and RTI, as well as suggest some future directions for the development of RTI based on prevention science. The chapter begins with the early history of the application of public health and prevention concepts to mental health and academic issues, and traces prevention efforts forward to the introduction of prevention science in the 1990s and some key developments in the last decade. Although it might seem logical to interweave developments in RTI with my description of developments in prevention science, the incorporation of prevention concepts into RTI has sometimes been asynchronous with advances in prevention. For example, prevention terminology that was abandoned in the 1980s can still be found in some contemporary RTI writings, alongside ideas and technology that also represent prevention science's cutting edge (e.g., Fuchs et al. 2012).

To reflect this asynchronous development of prevention and RTI, after an overview of the evolution of prevention concepts and technology, I briefly trace the history of RTI, starting with its origins in reading research, but also recognizing its influences from special education and school psychology. As a part of this chapter, the extent to which key practices within the current forms of RTI meet contemporary prevention science standards of evidence, and where more research is needed to support RTI practice is discussed. The chapter concludes by specifying areas where more empirical support for, and development of, RTI as a preventive intervention delivery system are needed.

A Brief History of Prevention and Prevention Science

Prevention has a long history, and its earliest applications in the USA were tied in with the prevention of illness and infectious disease by targeting infectious and environmental risks (Turnock 2012). This summary begins when contemporary prevention concepts began to emerge, shortly after World War II, and provides an overview of important developments in prevention that are relevant to RTI. A more comprehensive history can be obtained from other sources (e.g., Mrazek and Haggerty 1994; Susser and Stein 2009; Turnock 2012).

The Emergence of Comprehensive Prevention Models (1950–1970)

Following World War II, there was an increased interest in the application of public health and prevention principles to address the needs of a growing population, including returning veterans. A new conceptual framework, preventive medicine, was proposed which envisioned public health officials, private practitioners, and local organizations working together to promote the health of individuals and communities. Two leaders in this field were Leavell and Clark (1953; Clark and Leavell 1958), who outlined a model of prevention that divided prevention activities into two phases, with multiple levels within those phases. The two phases were distinguished by when the activities within them took place relative to the development of the disease targeted for prevention. Primary prevention activities took place while individuals were healthy, and were intended to prevent individuals from getting a disease. Secondary prevention activities took place early in a disease's course when symptoms were not obvious, and were meant to interrupt the disease process while it was latent. As Leavell and Clark's model made its way into wide use, primarily through its adoption by the Commission on Chronic Illness (1957), two aspects of the framework changed. The phases came to be known as levels, and a third level, tertiary prevention, was added (Mrazek and Haggerty 1994). The new level, tertiary prevention, described preventive activities that took place after a person had been diagnosed with a disease and were aimed at limiting the impact of the disease or facilitating rehabilitation. The model came to be known as "Leavell's levels" (Jekel et al. 2001).

Not only was there interest in the prevention of physical diseases at this time but also the prevention of mental illness and social deviancy. This focus was reflected in the formation of the Joint Commission on Mental Illness and Health as a part of the Mental Health Act of 1955. Organizers of this commission viewed schools as a promising setting for prevention activities (Ewalt 1957). Reading was an area of particular interest to prevention advocates, as reading difficulties were seen as important in and of themselves, and also as a risk factor for juvenile delinquency (Eisenberg 1962). In a foreshadowing of some key RTI practices, routine screening for reading disorders using oral reading passages was a suggested activity for pediatricians (Eisenberg 1959), and the use of brief remedial instruction to distinguish between children with reading problems due to neurological deficits versus poor instruction was suggested (Rabinovitch et al. 1956).

Among the mental health professionals interested in the application of prevention concepts to areas other than physical diseases was Gerald Caplan, a psychiatrist who had extensive experience treating children in orphanages in Israel following World War II. After attending lectures by Leavell while they both worked at Harvard University (Erchul and Schulte 1993), Caplan described how Leavell's three levels of prevention could be applied in community mental health efforts (Caplan 1964). As a part of this work, he also outlined a number of different strategies that could be employed by mental health professionals doing preventive work. One of his key prevention strategies was consultation, where mental health professionals worked with caregivers, such as teachers, to address problems the caregivers were encountering with their clients (Caplan 1970). Although his work had little focus on academic problems, Caplan's use of consultation became a well-accepted strategy for many other professionals' preventive efforts in schools (e.g., Meyers et al. 1979).

New Focus, New Models of Prevention (1971–1990)

Over time, as the focus of public health broadened beyond infectious diseases, it became clear that the conceptual framework and methods of prevention needed to be adapted. Germ theory, the underlying model for much of public health in the first half of the century, did not fit well with problems such as lung cancer and heart disease, which were now central concerns. A new view of disease came to predominate where an illness or disorder was viewed as the result of multiple causes acting in concert (Susser and Stein 2009).

With this new view, Leavell's levels presented problems as a model for preventive efforts. Its designations of primary, second, and tertiary prevention were based on knowing the disease mechanism and an individual's status relative to the course of a disease (Mrazek and Haggerty 1994). But chronic diseases often had multiple causes that were not fully understood, and disease status often fell on a continuum making it difficult to discern at what point a person "had" the disease. To reflect the changing perspective, Gordon (1983) proposed an alternate taxonomy that did not require an understanding of causes of disorders or the course of a disease, but classified prevention strategies based on the population groups to which they would be applied.

Gordon's (1983) system also had three levels, universal, selected, and indicated. Universal strategies were desirable for everyone. For example, seatbelts are a universal intervention strategy because they pose minimal risk and discomfort, and when used in the population at large, are likely to confer benefit to the group. Selective strategies were appropriate only for a subgroup of the population who had some characteristic that placed the subgroup particularly at risk. Generally, selected interventions have a lower benefit/risk ratio or are too costly to consider applying to the entire population. Avoidance of alcohol by pregnant women is a selective prevention strategy. Indicated strategies were appropriate only for persons who had a specific characteristic that placed them personally at risk, Strategies at this third level have higher costs and/or risks, such that individual evaluation is required to decide whether the strategy should be applied. For example, medication to lower blood pressure is an indicated strategy for the prevention of a heart attack or stroke in a person with hypertension.

Leavell and Clark's (1953) and Gordon's (1983) classification systems were not intended to be interchangeable. Although universal interventions in Gordon's framework and primary prevention strategies in Leavell and Clark's framework align well, secondary prevention in the Leavell framework encompasses Gordon's selected and indicated interventions. Gordon's model explicitly excluded intervention strategies directed at persons with a diagnosed disorder (Mrazek and Haggerty 1994), which was the focus of tertiary intervention in the Leavell and Clark system.

Within the area of epidemiology, this shift to a focus on chronic diseases also had implications. Surveillance, meaning the monitoring of disease in populations, has always been a key part of epidemiology and public health (Jekel et al. 2001). However, during this era, targets for surveillance in public health were broadened beyond communicable diseases to include aspects of health such as child growth and development and mental illness (Declich and Carter 1994).

Also changing within epidemiology was the analysis of risk factors. The new focus on chronic disease required analytic strategies that could address multiple risk factors and interactions between them, and trace the development of diseases and conditions across time. Sophisticated research designs (e.g., prospective cohort and case control studies) became more common, and epidemiology emerged as a distinct discipline with a prominent role in prevention (Susser and Stein 2009).

This shift in focus also paved the way for many advances in the application of prevention concepts to a broader range of human problems. Although prevention concepts had been applied to mental health and education previously, efforts in this area had been sporadic and lacking focus, partly because of the poor understanding of how risk factors operated in psychiatric disorders (Mrazek and Haggerty 1994). Better definition and measurement of psychiatric disorders, as well as an appreciation of the complexities involved in understanding the interplay of risk and protective factors at multiple levels (genetic, biological, individual, family, systems), led to significant funding for prevention research in education and mental health in the 1980s. For example, in response to concerns about the high prevalence of reading problems and advances taking place at that time in research on predictors of reading problems, the 1985 Health Research Extension Act authorized a new and expanded program of research on reading difficulties through the National Institute of Child Health and Human Development (NICHD).

The Emergence and Growth of Prevention Science (1991–present)

The increased funding for broader prevention work in the 1980s, as well as recognition that the greatest areas of need in prevention for children had changed to addressing psychosocial risks (Baumeister et al. 1990), led to a concerted effort to increase the rigor of prevention efforts focused on psychosocial risks and disorders. As a part of these efforts, the field of prevention science emerged in the early 1990s (Coie et al. 1993). Although definitions of prevention science vary, common aspects in the definitions of the field are that it: (a) is interdisciplinary; (b) is framed within an ecological/developmental/epidemiological perspective that emphasizes multiple nested systems, age-related patterns of competence and difficulties, and a population-based consideration of risk factors; (c) emphasizes scientific rigor; and (d) characterizes prevention work as a research cycle that proceeds through distinct phases, each with their own standards of evidence (Coie et al. 1993; Herman et al. 2012; Kellam et al. 1999; Kellam and Langevin 2003; National Research Council and Institute of Medicine 2009; Stormont et al. 2010). The mission of prevention science involves "identifying specific antecedents of later mental disorders or mental health and targeting these risk or protective factors to inhibit or enhance their influence on the life course" (Kellam et al. 1999, p. 466). Elements of prevention retained in prevention science from earlier eras included Gordon's (1983) tiered model for classifying preventive interventions, and the focus on epidemiological approaches to screening and surveillance.

The articulation of the preventive intervention research cycle was one of the most important features distinguishing prevention science from earlier prevention work (Kellam et al. 1999). The research cycle was widely disseminated in an Institute of Medicine Report, *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research* (Mrazek and Haggerty 1994). The cycle begins with the identification of a problem or disorder, a review of information on its prevalence, and a search of existing research to identify risk factors that precede the disorder. A theory of the problem is then developed that links the risk factor to the disorder. The theory provides both a target for preventive intervention, and an explanation of the process by which the risk factor leads to the eventual disorder, and how the process might be interrupted with the intervention. Studies to test the theory and validate the preventive intervention begin with pilot studies, followed by efficacy studies testing the intervention's effect under ideal conditions. When efficacy trial results are promising, effectiveness studies examining the intervention's effect under real-world conditions are conducted. The final stage in the preventive intervention research cycle is large-scale implementation of preventive interventions.

The dual emphasis on validating a causal theory and demonstrating a preventive intervention's impact in prevention science logically leads to a focus on causal experimental designs rather than correlational ones, particularly in efficacy and effectiveness trials. It also requires that the researcher demonstrate that the intervention exerted its preventive effect on the targeted disorder through a change in the risk factor targeted (Kellam et al. 1999). In this way, the preventive intervention trial contributes information about whether the underlying theory of the cause of the disorder is correct, or needs to be modified.

The preventive science framework has evolved since its introduction. In 2005, the Society for Prevention Research (SPR) issued a set of standards that specifying criteria for designating preventive interventions and policies as evidence based. The document outlined three sets of overlapping standards that corresponded to three of the phases of prevention research: efficacy trials, effectiveness trials, and dissemination (Flay et al. 2005). For example, for a prevention strategy to be considered to have met efficacy standards, it is to have been tested in at least two field trials, showed consistent positive effects, and had at least one positive effect that was significant at follow up. To be considered effective, the developers of the same intervention strategy would have to demonstrate that manuals or technical support for the preventive intervention were available, the intervention had been evaluated in real-world conditions, and that the intervention outcomes were practically important. Finally, an intervention meeting standards for dissemination would have evidence that it can be "scaled up" to widespread implementation while maintaining positive effects; clear information regarding costs such as investments in staff training, time, and materials; and evaluation tools available to assist providers in determining whether the program was delivered as intended and what outcomes were obtained.

As noted in a recent follow-up report, the articulation of this research cycle in Mrazek and Haggerty (1994), along with multiple pieces of legislation at that time that began requiring federal programs to supply evidence of their effectiveness, resulted in a marked increase in randomized, controlled trials of preventive interventions (National Research Council & Institute of Medicine 2009). In turn, these efforts produced evidence that several prevention programs achieved their aims. Among the successful preventive programs with a school component highlighted in the document were Fast Track (Conduct Problems Prevention Research Group 2007) and the Seattle Social Development Project (Hawkins et al. 2008).

However, the report and several other documents also noted that there were underdeveloped portions of prevention science technology, including a lack of agreed upon guidelines for making decisions based on screening results, and very limited models and evidence about how to move successful preventive interventions into widespread practice (National Research Council & Institute of Medicine 2009; Rohrbach et al.

2006). Since the introduction of the prevention research cycle, the dissemination phase has been further subdivided into dissemination, adoption, implementation, and maintenance or sustainability (SPR MAPS II Task Force 2008). Research on these areas has been labeled “type 2 translational research,” where type 1 translational research is bringing basic science to bear to improve health practices or moving from “bench to bedside” (SPR MAPS II Task Force 2008).

Today, these areas are the frontiers of prevention science where development is needed. Additional areas in need of future research are balancing the need for program fidelity with the need for appropriate adaptations to assure an intervention works across different sites or cultural groups, and assessing the implications of advances in genetics, epigenetics, and assessment for developing a “personalized” rather than population or group-based assessment of risk (Castro et al. 2005; National Research Council & Institute of Medicine 2009).

Prevention and RTI

Prevention has exerted influence on RTI in multiple ways. First, funding for prevention research, notably from the National Institute of Child Health and Development (NICHD), was instrumental in producing the research foundation for the earliest iterations of RTI, when its focus was exclusively on better screening and diagnosis of reading disorders. Second, practices that developed within prevention, such as consultation, made their way into school-based practices that were folded into RTI as its function broadened. Finally, some concepts or practices from prevention made their way directly into RTI, such as the practice of universal screening.

NICHD Research on Reading Disabilities

As noted earlier in this chapter, there has been a longstanding interest in the reading disabilities in the prevention literature, both as a disorder with high prevalence and as a risk factor for the development of other disorders (Eisenberg 1962). In the early 1980s, NICHD sponsored a number of key studies regarding reading disabilities. One of these studies was the Connecticut Longitudinal Study, which followed a representative sample of 445 children entering public school in Connecticut for multiple years, starting in 1983 (Shaywitz et al. 1990). Because the children were a representative sample of the entire school population and the study was longitudinal, different definitions of reading disability could be examined as well as risk factors in the development of reading disorders. Findings from this study were instrumental in casting doubt on the validity of the ability–achievement discrepancy method of diagnosing reading disorders and also in demonstrating that early difficulties with phonemic processing were predictive of later difficulties in reading (Fletcher et al. 1994, 1998).

A few years later, in 1985, the Health Research Extension Act charged NICHD with the task of improving the quality of reading research by conducting long term, multidisciplinary research (Grossen 1997). This new initiative provided the funding for more than 100 researchers at 14 centers, and much of the early RTI research arose from this source of funding, including the Vellutino et al. (1996) study cited in the introduction. In 2000, Torgesen, another researcher funded by NICHD, summarized the findings from five methodologically rigorous studies concerning the prevention of reading disabilities. Based on the results from these studies, he estimated that by using appropriate instructional methods to help children acquire initial word reading skills, early reading achievement for at least 50 % of children who were at risk for reading disorders could be normalized. In keeping with the longitudinal focus of prevention research, he also noted that

> Of course, we are going to follow the children’s development through the next several years of elementary school to examine their long-term reading development. Without such long-term follow up, we will not really know whether we were able to “prevent” reading problems in these children.… (Torgesen 2000, p. 62)

Torgesen also noted problems with false positives in the five studies, that is, a high rate of children were identified as at risk in these studies who would not later develop reading disabilities and were inappropriately placed in preventive services.

The Problem-Solving Model

The mounting evidence from prevention studies in reading, such as those summarized by Torgesen (2000) as well as longstanding concerns about the quality of special education, influenced several key pieces of federal legislation under consideration at that time. One of these was the No Child Left Behind Act of 2001 (NCLB 2002), which included the Reading First and Early Reading First initiatives. These NCLB initiatives provided funding to incorporate research-based findings on reading into general education instruction (Lyon et al. 2005). A few years later, the 2004 reauthorization of the Individuals with Disabilities Education Improvement Act (U. S. Department of Education 2004) also referenced the emerging reading disorders prevention research and explicitly authorized the use of RTI as a means of diagnosing reading disabilities. These changes set the stage for large-scale adoption of RTI by states.

However, data from a key stage of the preventive intervention research cycle were missing. Although there was a body of research studies showing that high-quality early instruction reduced reading problems in the lower grades, and that children's response to instruction was a useful diagnostic criterion for learning disability diagnosis, in almost all of the studies, the early reading interventions had been implemented by researcher-trained tutors (e.g., Vaughn et al. 2003; Vellutino et al. 1996). Despite the promise of RTI procedures in research, effectiveness studies and evidence that RTI could be successfully disseminated and implemented by school staff were not available.

But another set of practices that offered a good fit with the research-based models of RTI had already been scaled up to school, district, and state levels—problem-solving consultation. Problem-solving consultation as a service delivery model had been developed a decade and a half earlier in response to concerns about large numbers of children referred and placed in special education after the passage of the Education for All Handicapped Children Act of 1975. It was an adaptation of Bergan's (1977) behavioral consultation model, developed as a preventive intervention for children in Headstart transitioning into school. In turn, Bergan's model had been influenced by even earlier prevention work, Caplan's (1970) consultation model (Bergan 1977; Bergan and Tombari 1976).

Different models of problem-solving consultation had been adopted and widely implemented in several states well before the introduction of RTI (e.g., Graden et al. 1985; Ikeda et al. 1996; Telzrow et al. 2000). For example, in Iowa, Bergan's model was introduced in the early 1980s (Gresham 1982) and a few years later included as part of a project funded by the US Department of Education, the Relevant Education Assessment and Interventions Model (RE-AIM: Ikeda et al. 1996). One iteration of this approach in Iowa came to be known as the "Heartland model" which consisted of problem-solving consultation, progress monitoring using curriculum-based measurement, and formation of building assistance teams (Grimes and Kurns 2003; Ikeda et al. 1996).

As it became evident that these problem-solving-based prereferral intervention models had considerable overlap with the RTI models proposed by reading researchers, and that the problem-solving consultation models had already gone to scale, they were recognized as a second variant of RTI (Berkeley et al. 2009; Fuchs et al. 2003). Two major differences between the two variants were (a) the use of a standard intervention protocol in versus an individualized approach to designing interventions and (b) the more extensive and methodologically rigorous research base supporting the standard protocol RTI variant (Fuchs et al. 2003; Fuchs et al. 2012).

Other Prevention Concepts in RTI

As the previous two sections have illustrated, prevention funding and early preventive intervention models both influenced the development of RTI. However, a number of prevention concepts have simply made their way directly into RTI. For example, contemporary RTI is characterized as a multi-tiered preventive approach, where increasingly intensive interventions are applied to smaller numbers of students. Variants of both Leavell's levels (Leavell and Clark 1953) and Gordon's (1983) framework can be found in RTI writing. For example, Walker et al. (1996) used Leavell's primary, secondary, and tertiary prevention levels to describe a continuum of behavior support for students in schools. These levels were eventually incorporated into their school-wide positive behavior support model (Horner et al. 2009). Leavell's levels are also found in other multi-tiered RTI models (e.g., Fuchs et al. 2012; Malecki and Demaray 2007). The influence of Gordon's multi-tiered (1983) prevention framework is evident in Good, Simmons, Kame'enui, and Wallin's (2002) three-level system for screening and classifying beginning readers' need for interventions (benchmark, strategic, and intensive), and these levels are also found in the RTI model described by Burns and Gibbons (2008).

Finally, universal screening has been called "RTI's greatest accomplishment" (Fuchs and Vaughn 2012). As a part of RTI, routine screening of all children for reading and mathematics difficulties is an increasingly accepted practice. Although many of the measures used for universal screening in RTI grew out of special education research, such as curriculum-based measures (Deno and Mirkin 1979; Fuchs et al. 1984), their use in universal screening reflects the key concept of surveillance in prevention (Herman et al. 2012). Furthermore, the most appropriate statistics to establish the validity of these measures for screening draw directly from epidemiological statistical methods and screening theory (Jekel et al. 2001; Susser and Stein 2009).

Where Are We Now? Prevention Science and RTI

The previous sections have described the multiple strands of influence that prevention work has had upon RTI. Now that RTI is recognized in federal policy, has been implemented in at least 15 states, and is under development in most of the remaining states (Berkeley et al. 2009), it is useful to examine where RTI stands in relation to current practice in prevention science. As indicated earlier, one of the recent advances in prevention has been articulation of a research cycle for prevention research and the articulation of standards of evidence by the SPR that specify what types of evidence should be present from three stages of the research cycle (efficacy, effectiveness, and dissemination) *before* a practice becomes policy (Flay et al. 2005). In this section, contemporary RTI is examined in relation to these standards. Although a review of contemporary RTI in relation to all 47 of the SPR standards is beyond the scope of this chapter, aspects of RTI relevant to salient components of these standards at each of these three levels of evidence are discussed.

Efficacy

Among the key SPR evidence standards for considering a policy, practice, or intervention to be efficacious are the presence of a clear and specific statement of efficacy that describes the prevention practice, the outcome it produces, and the population to which it is applied; a description of the intervention that permits replication; and evidence that the practice produced statistically significant effects in studies using research designs that permit unambiguous causal statements.

The strongest support for many of the practices in RTI models is at this phase of the preventive intervention research cycle. There are multiple randomized, controlled studies demonstrating that explicit instruction in five key areas (i.e., phonemic awareness, phonics, fluency, vocabulary, and text comprehension) prevent early reading failure (NICHD 2000; Torgesen et al. 2000). There is also strong evidence for a number of the key practices that are part of RTI models for preventing social and behavior problems, such as modifying the classroom environment to decrease problem behaviors and teaching students desired skills and behaviors that replace the undesirable behaviors (Institute of Education Sciences 2008). However, RTI was expanded to areas of academic functioning where the evidence of its preventive benefit, or the efficacy of many components, was scant or absent before it became policy, such as mathematics and reading intervention for upper grades (Gersten et al. 2009; Vaughn et al. 2010).

The SPR standards also require significant effects for at least one outcome on long-term follow-up. Here, as Torgesen noted in 2000 in the quote provided earlier, the evidence regarding the impact of RTI on preventing reading problems in later grades is lacking. There are now indications that the screening procedures used in RTI overestimate children's risk for reading problems in later grades. As a result, too many children are directed into secondary prevention efforts that they do not need (Fuchs et al. 2012).

Effectiveness

Among the key SPR evidence standards for considering a policy, practice, or intervention to be effective are (a) the availability of methodologically strong studies showing a positive impact of the practice when implemented under real-world conditions for the target population to which the practice is intended to generalize, and (b) positive outcomes in a long-term follow-up of participants from one or more effectiveness studies.

As noted by Denton (2012) and VanDerHeyden et al. (2012), little existing research has examined RTI as an integrated whole, which is how the model has evolved in recent years and is implemented in schools (Berkeley et al. 2009). Examining the impact of RTI as a whole on the diverse body of students likely to be affected by it when it is implemented by school staff is required to demonstrate its effectiveness. Noting this lack of evidence, Hughes and Dexter (2008) searched for and then reviewed field studies of RTI to find evidence of its effectiveness. They found 16 studies examining integrated models of RTI as a means of addressing children's academic or behavioral concerns. Fourteen of the studies had teachers implementing at least some of the tiers. Although results from these studies were generally positive, almost all of them employed quasi-experimental or descriptive designs. Since the Hughes and Dexter review, additional studies of RTI models have been conducted in real-world settings with positive results (e.g., VanDerHeyden et al. 2012; Vaughn et al. 2009; Simmons et al. 2011). Although long-term follow-up studies remain sparse, manuals and technical support materials for RTI are increasingly available (e.g., Burns and Gibbons 2008).

Dissemination

Among the key SPR evidence standards for considering a policy, practice, or intervention to be ready for dissemination are prior evidence of efficacy and effectiveness, availability of monitoring and evaluation tools for providers, and a demonstrated practical public health impact of the practice when it is implemented in the field.

As has been noted in earlier sections, RTI was incorporated into federal legislation and practice well before there was a substantial body of evidence supporting its effectiveness, although research in this area is emerging. Furthermore, as noted by others (Fuchs et al. 2003; Fuchs et al. 2010), the model of RTI that has been widely disseminated and implemented, problem-solving, lacks strong evidence to support its efficacy or effectiveness. However, it does provide what many of the more standard protocol variants of RTI lack, monitoring and evaluation tools that help facilitate implementation by practitioners (e.g., Burns and Gibbons 2008).

The discrepancy between the SPR standards of evidence and the incorporation of RTI into federal and state policy may suggest that the prevention science movement and its focus on evidence-based intervention has had limited impact. However, it is important to note that the small, but robust body of efficacy studies, and the limited effectiveness data that were available for RTI before it was implemented, actually represent more evidence than has been available for many other special education practices prior to their implementation (Gersten and Dimino 2006).

Although the SPR evidence standards call for monitoring and evaluation tools for providers and data on RTI's public health impact before implementation, these two important aspects of sustaining and institutionalizing RTI are just emerging. At present, there are six regional resource centers that support effective implementation of RTI (Berkeley et al. 2009). There are also early studies of large-scale implementation of RTI that support its continued implementation, but also suggest that refinements are necessary.

Among recent findings relevant to RTI dissemination and implementation are the results from state-wide implementation of academic or behavioral RTI models. Horner et al. (2009) reported data from a two-state implementation of school-wide positive behavior support, indicating that schools where the model was implemented were perceived as safer and the intervention also may have had an impact on academic outcomes. In another study, Torgesen (2009) provided results from the large-scale implementation of RTI in Reading First schools in Florida to prevent reading problems. Over 3 years of implementation, he reported dramatic drops in referral rates of students for learning disabilities services, with percentages of students identified as learning disabled dropping by more than half in grades kindergarten through second, and about 40 % in third grade. However, the actual reading outcomes for students did not improve to the same extent. Using the percent of students scoring at or below the 20th percentile as an indication of the rate of students with significant reading problems in the early grades, Torgesen reported that by the third year of implementation, the percent of students scoring below this cutoff had dropped from first-year highs of 23–27 % to 14–19 %.

There are also some interesting findings emerging from continuing research that suggest that some aspects of RTI are in need of modification. In two recent articles appraising the current status of RTI (Fuchs et al. 2012; Fuchs and Vaughn 2012), concerns were raised about the high rate of false negatives in RTI screening, the use of slopes as a predictive tool for appraising risk, and evidence suggesting requiring a second tier of intervention before special education for some children was unnecessary based on their risk profile and delayed their access to much needed services.

Prevention Science and Future Directions for RTI Research and Practice

As this chapter has documented, both prevention and RTI are "living" bodies of knowledge that have evolved over time. Throughout its short history, RTI has benefited from concepts and techniques drawn directly and indirectly from prevention. Although RTI falls short of current standards for evidence-based prevention science policy and practice, it also represents a significant step forward in the use of research data to design, implement, and evaluate school-based practices.

One of the most critical research needs identified in prevention science is for type 2 translational research that provides us with better models of large-scale implementation and knowledge about the factors critical in this process (SPR MAPS II Task Force 2008). It is also this phase of the research cycle where much work is needed in RTI. As such, it is likely that RTI will benefit from work on this topic in prevention science. However, there is also a strong likelihood of reciprocal benefit—RTI is one of the first policies born in an era where there has been a concerted, planful effort to incorporate research-based findings into policy. As such, there are many lessons to be learned for prevention scientists from the emerging findings about RTI implementation. These lessons are likely to enhance RTI, but also will enhance policy development in multiple areas important to improving outcomes for children.

References

Barnes, A. C., & Harlacher, J. E. (2008). Clearing the confusion: Response-to-intervention as a set of principles. *Education and Treatment of Children, 31,* 417–431.

Baumeister, A. A., Kupstas, F., & Klindworth, L. M. (1990). New morbidity: Implications for prevention of children's disabilities. *Exceptionality, 1,* 1–16. doi:10.1080/09362839009524738.

Bergan, J. R. (1977). *Behavioral consultation*. Columbus: Merrill.

Bergan, J. R., & Tombari, M. L. (1976). Consultant skill and efficiency and the outcomes of consultation. *Journal of School Psychology, 14,* 3–14.

Berkeley, S., Bender, W. N., Gregg, L. P., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42,* 85–95. doi:10.1177/0022219408326214.

Burns, M. B., & Gibbons, K. A. (2008). *Implementing response-to-intervention in elementary and secondary schools*. New York: Routledge.

Caplan, G. (1964). *Principles of preventive psychiatry*. New York: Basic Books.

Caplan, G. (1970). *The theory and practice of mental health consultation*. New York: Basic Books.

Clark, E. G., & Leavell, H. R. (1958). Levels of application of preventive medicine. In H. R. Leavell & E. G. Clark (Eds.), *Preventive medicine for the doctor in his community* (pp. 13–39). New York: McGraw-Hill.

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., et al. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist, 48,* 1013–1022.

Conduct Problems Prevention Research Group. (2007). Fast track randomized controlled trial to prevent externalizing psychiatric disorders: Findings from grades 3 to 9. *Journal of the American Academy of Child & Adolescent Psychiatry, 46,* 1250–1262.

Declich, S., & Carter, A. O. (1994). Public health surveillance: Historical origins, methods and evaluation. *Bulletin of the World Health Organization, 72,* 285–304.

Deno, S., & Mirkin, P. K. (1977). *Data-based program modification*. Minneapolis: University of Minnesota.

Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities, 45,* 232–243.

Eisenberg, L. (1959). Evaluation of specific reading disability in children. *Pediatrics, 23,* 997–1003.

Eisenberg, L. (1962). Preventive psychiatry. *Annual Review of Medicine, 13,* 343–360.

Erchul, W. P., & Schulte, A. C. (1993). Gerald Caplan's contributions to professional psychology: Conceptual underpinnings. In W. P. Erchul (Ed.), *Consultation in community, school, and organizational practice* (pp. 3–40). Washington, DC: Taylor & Francis.

Ewalt, J. R. (1957). Goals of the Joint Commission on Mental Illness and Health. *American Journal of Public Health, 47,* 19–24.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6,* 151–175.

Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities: Research and Practice, 13,* 186–203.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Fowler, A., et al. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 1–18.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21,* 449–460.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18,* 151–171.

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76,* 301–323.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78,* 263–279.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13,* 204–219.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-Intervention: A decade later. *Journal of Learning Disabilities, 45,* 195–203. doi:10.1177/0022219412442150.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., et al. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009–4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Gersten, R., & Dimino, J. A. (2006). RTI (Response to Intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41,* 99–108. doi:10.1598/RRQ.41.1.5.

Good, R. H., Simmons, D., & Kame'enui, E. (2002). *Summary of decision rules for intensive, strategic, and* *benchmark instructional recommendations in kindergarten through third grade (Technical Report No. 11)*. Eugene: University of Oregon.

Gordon, R. S. Jr. (1983). An operational classification of disease prevention. *Public Health Reports, 98,* 107–109.

Graden, J. L., Casey, A., & Christenson, S. L. (1985). Inplementing a prereferral intervention system: Part I. *The model, 51*, 377–384.

Gresham, F. (1982). *Handbook for behavioral consultation*. Des Moines: State of Iowa Department of Public Instruction.

Grimes, J., & Kurns, S. (2003, December). *An intervention-based system for addressing NCLB and IDEA expectations: A multiple tiered model to ensure every child learns*. Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-intervention symposium, Kansas City.

Grossen, B. (1997). 30 years of research: What we now know about how children learn to read. Santa Cruz: Center for the Future of Teaching and Learning.

Hawkins, J. D., Brown, E. C., Oesterle, S., Arthur, M. W., Abbott, R. D., & Catalano, R. F. (2008). Early effects of Communities That Care on targeted risks and initiation of delinquent behavior and substance use. *Journal of Adolescent Health, 43,* 15–22.

Herman, K. C., Riley-Tillman, T. C., & Reinke, W. M. (2012). The role of assessment in a prevention science framework. *School Psychology Review, 41,* 306–314.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., et al. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*, 133–144. doi:10.1177/1098300709332067.

Hughes, C. A., & Dexter, D. D. (2008). Field studies of RTI programs, revised. http://www.rtinetwork.org/learn/research/field-studies-rti-programs.

Ikeda, M. J., Tilly, W. D. III, Stumme, J., Volmer, L., & Allison, R. (1996). Agency-wide implementation of problem solving consultation: Foundations, current implementation, and future directions. *School Psychology Quarterly, 11,* 228–243.

Jekel, J. F., Katz, D. L., & Elmore, J. G. (2001). *Epidemiology, biostatistics, and preventive medicine* (2nd ed.). Philadelphia: W. B. Saunders.

Kellam, S. G., Koretz, D., & Mościcki, E. K. (1999). Core elements of developmental epidemiologically based prevention research. *American Journal of Community Psychology, 27,* 463–482. doi:10.1023/A:1022129127298.

Kellam, S. G., & Langevin, D. J. (2003). A framework for understanding "evidence" in prevention research and programs. *Prevention Science, 4,* 137–153. doi:10.1023/A:1024693321963.

Leavell, H. R., & Clark, E. G. (1953). *Textbook of preventive medicine*. New York: McGraw-Hill.

Lyon, G. R. (1994). Methodological issues and strategies for assessing developmental change and evaluating response to intervention. In D. L. Speece & B. K. Keogh (Eds.), *Research on classroom ecologies* (pp. 213–227). Mahwah: Erlbaum.

Lyon, G. R., Shaywitz, S. E., Shaywitz, B. A., & Chhabra, V. (2005). Evidence-based reading policy in the United States: How scientific research informs instructional practices. In D. Ravitch (Ed.), *Brookings papers on education policy* (Vol. 2005, pp. 209–250). Washington, DC: Brookings Institution Press. doi:10.1353/pep.2005.0009.

Malecki, C. K., & Demaray, M. K. (2007). Social behavior assessment and response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 161–171). New York: Springer.

Meyers, J., Parsons, R. D., & Martin, R. (1979). *Mental health consultation in the schools*. San Francisco: Jossey-Bass.

Mrazek, P. J., & Haggerty, R. J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers of preventive intervention research*. Washington, DC: National Academy Press.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Research Council and Institute of Medicine. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.

No Child Left Behind (NCLB) Act of 2001, Pub L. No. 107–110, § 115, Stat. 1425 (2002).

Rabinovitch, R. D., Drew, A. L., Dejong, R. N., Ingram, W., & Withy, L. (1956). A research approach to reading retardation. In R. Mcintosh (Ed.), *Neurology and Psychiatry in Childhood*. Baltimore: Williams & Wilkins.

Rohrbach, L. A., Grana, R., Sussman, S., & Valente, T. W. (2006). Type II translation: Transporting prevention intervention research to real-world settings. *Evaluation & the Health Professions, 29,* 302–333.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association, 264,* 998–1002.

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Oi-man, K., Simmons, L., Johnson, C., et al. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergartners' response. *Exceptional Children, 77,* 207–228.

Society for Prevention Research MAPS II Task Force. (2008). *Type 2 translational research: Overview and definitions*. Fairfax.

Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disabilities. *Journal of Educational Psychology, 94,* 735–749.

Stormont, M., Reinke, W. M., & Herman, K. C. (2010). Introduction to the special issue: Using prevention science to address mental health issues in schools. *Psychology in the Schools, 47,* 1–4. doi:10.1002/pits.20447.

Susser, M., & Stein, Z. (2009). *Eras in epidemiology*. Cary: Oxford University Press.

Telzrow, C. F., McNamara, K., & Hollinger, C. L. (2000). Fidelity of problem-solving implementation and relationship to student performance. *School Psychology Review, 29,* 443–461.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., & Conway, T. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 1999,* 579–593.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15,* 55–64.

Torgesen, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in Reading First schools. *Child Development Perspectives, 3,* 38–40. doi:10.1111/j.1750-8606.2009.00073.x.

Turnock, B. J. (2012). *Public health: What it is and how it works*. Burlington: Jones & Bartlett Learning.

U.S. Department of Education. (2004). Individuals with Disabilities Improvement Act of 2004, Pub. L. 108–466. *Federal Register, 20,* 35802–35803.

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Educational Research Journal, 49,* 1251–1284.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75,* 165–183.

Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. A., et al. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 39,* 3–21.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., et al. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4,* 194–209.

Problem-Solving Consultation

William P. Erchul and Caryn S. Ward

> Problem solving is the essence of consultation (Zins and Erchul 2002, p. 631).
>
> It is apparent that a PS [problem-solving]/RTI [response-to-intervention] model is not only consistent with, but in fact dependent on, the delivery of effective consultation services… for successful implementation (Gutkin and Curtis 2009, p. 594).
>
> [T]he problem-solving approach used in RTI described in the behavioral consultation literature… can represent a comprehensive framework to RTI (Kratochwill et al. 2007, p. 27).

These three quotes serve as a useful starting point because they illustrate the intertwined nature of problem-solving, consultation, and response to intervention (RTI)—the principal focus of this chapter.

Problem-solving represents a predictable sequence of events in which an issue of concern is identified, then defined and further analyzed to the point an appropriate intervention can be devised/selected and subsequently implemented to address the problem. The effect of the intervention on the problem is then assessed (Erchul and Martens 2010). Within education and psychology, the components of problem-solving may be specified through a series of questions:

1. Is there a problem and what is it?
2. Why is the problem happening?
3. What can be done about the problem?
4. Did the intervention work? (Tilly 2008, p. 18)

Consultation is an established vehicle for conducting problem-solving in schools. Although many definitions exist, consultation may be defined as "a method of providing preventively oriented psychological and educational services in which consultants and consultees form cooperative partnerships and engage in a reciprocal, systematic problem-solving process guided by ecobehavioral principles. The goal is to enhance and empower consultee systems, thereby promoting clients' well-being and performance" (Zins and Erchul 2002, p. 626). More specifically, *school consultation* is "a process for providing psychological and educational services in which a specialist (consultant) works cooperatively with a staff member (consultee) to improve the learning and adjustment of a student (client) or group of students. During face-to-face interactions, the consultant helps the consultee through systematic problem-solving, social influence, and professional support. In turn, the consultee helps the client(s) through selecting and implementing effective school-based interventions. In all cases, school consultation serves a remedial function and has the potential to serve a preventive function" (Erchul and Martens 2010, pp. 12–13).

The research base that undergirds consultation in schools has grown over time despite inherent difficulties in undertaking such investigations (Erchul and Sheridan 2014). For instance, because a consultant attempts to effect change in a third party (i.e., client) by working directly with

W. P. Erchul (✉)
Arizona State University, Tempe, AZ, USA
e-mail: William.Erchul@asu.edu

C. S. Ward
University of North Carolina, Chapel Hill, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_6

a second party (i.e., consultee), it is often difficult to attribute client change to a consultant's actions. Despite these shortcomings, meta-analytic and literature reviews of consultation outcomes (e.g., Lewis and Newcomer 2002; Reddy et al. 2000) provide support for consultation's effectiveness. Sheridan et al. (1996), for example, analyzed 46 school consultation studies published between 1985 and 1995. They found that consultation produced positive results in 67% of the studies reviewed, while 28% were neutral and the remaining 5% were negative. Busse et al. (1995) adapted meta-analytic procedures for single-case designs and documented an average client effect size of 0.95. More recently, randomized control trials of consultation have been conducted (e.g., Fabiano et al. 2010; Sheridan et al. 2012) and these methodologically superior studies hold great promise for advancing consultation as an evidence-based practice in schools, particularly in the contemporary era of RTI (Erchul and Sheridan 2014).

RTI is a data-driven process of implementing instruction and interventions to ensure that all students are successful academically and behaviorally (Batsche et al. 2005; Reschly and Bergstrom 2009; VanDerHeyden et al. 2007). Increasingly intensive instruction and interventions are provided within a tiered system of supports and structures. Although schools may utilize different models of RTI, most use a three-tiered system (Reschly and Bergstrom 2009). Within the first tier, all students receive high-quality research-based core instruction and behavioral supports provided by the classroom teacher(s). Students who are "at risk" are identified using universal screening assessments and other data. Increasing intensive instructional and behavioral interventions are implemented to assist students who are not making adequate progress in response to the high quality instruction within the secondary and tertiary tiers. Data collected through progress monitoring and formative assessments are utilized to evaluate instructional practices and facilitate educational decision-making processes. Throughout, teams are engaged in a problem-solving process to create systems of supports for students. The problem-solving process typically encompasses the four-step model of problem identification, problem analysis, plan generation and implementation, and plan evaluation (Kratochwill et al. 2007). In practice, RTI has been most commonly utilized to drive academic supports and resources; less emphasis has been placed on behavioral supports, which are most commonly addressed in practice by the implementation of positive behavior Intervention supports (PBIS), which is also a tiered system of supports (Horner et al. 2009). A more recent term used in practice and the professional literature is *multi-tiered systems of support* (MTSS; Sugai and Horner 2009). MTSS is an integrated streamlined system of supports, which encompasses all of the critical components of RTI and PBIS.

Reschly and Bergstrom (2009) indicated that, among other domains, *consultation methods for problem-solving* comprise an important research foundation for RTI/MTSS. Therefore, the primary purpose of this chapter is to: (a) describe school-based problem-solving consultation and aspects of RTI/MTSS and (b) examine their relationship to one another. The chapter concludes with a consideration of future research topics and implications for practice.

Problem-Solving Consultation

Origins and Evolution

As noted, the notion of problem-solving has always been intrinsic to consultation. Along these lines, D'Zurilla and Goldfried (1971) offered an early version of consultation that embeds behavior modification procedures within a problem-solving framework. They observed that problem-solving is a process that offers many possible alternatives, thereby increasing the likelihood individuals will select the most effective one available. Their model contains five problem-solving stages:

1. *General orientation:* Developing attitudes (e.g., acceptance) that problems occur in life, recognizing these problems, and inhibiting a tendency to act impulsively or not at all

2. *Problem definition and formulation:* Defining all dimensions of the problem in operational terms and identifying aspects relevant to the situation
3. *Generation of alternatives:* Brainstorming and combining various options for problem resolution
4. *Decision-making:* Forecasting the likely outcomes of each option
5. *Verification:* Assessing effectiveness by comparing obtained outcomes to predicted outcomes

Tharp and Wetzel (1969) similarly proposed a detailed process of applying principles of behavior modification in human service settings such as schools. Their approach recognizes the significant role direct care providers (e.g., parents, teachers) play serving as behavior change agents in natural settings. They identified three key participants: (a) *consultant,* an individual with expertise in behavior analysis; (b) *mediator,* an individual who controls reinforcers and can supply them contingently based on a target's behavior; and (c) *target,* an individual with a problem that can be addressed through behavioral techniques. In today's consultation literature, *consultee* is of course synonymous with *mediator,* and *client* with *target*.

Extending the ideas advanced by D'Zurilla and Goldfried (1971) and Tharp and Wetzel (1969), John Bergan advanced his model of behavioral consultation (Bergan 1977; Bergan and Kratochwill 1990). Bergan's model reflects a four-stage problem-solving process, and its well-known stages include three separate interviews, each of which contains specific goals and objectives that the consultant is expected to address (i.e., problem identification, problem analysis, plan implementation, problem evaluation). Plan implementation more recently has been termed *treatment implementation* and problem evaluation has been termed *treatment evaluation* (Sheridan and Kratochwill 2008) or *plan evaluation* (Kratochwill et al. 2002). These slight name changes notwithstanding, the key stages of behavioral consultation have remained essentially the same for nearly 40 years (Erchul and Schulte 2009).

Some classic behavioral consultation research findings relating to consultant interviewing skills include: (a) defining the problem in behavioral terms is the best predictor of ultimate problem resolution (Bergan and Tombari 1975, 1976); (b) using behavioral terminology (vs. medical model cues) leads to higher expectations by teachers about their ability to teach children with academic problems (Tombari and Bergan 1978); and (c) asking a consultee to identify the resources needed to implement an intervention plan—rather than telling him or her what to do—results in odds 14 times higher that a consultee will actually identify these resources (Bergan and Neumann 1980).

In the intervening years since Bergan's behavioral consultation model was introduced and the research cited in the previous paragraph was conducted, the practice of school-based problem-solving consultation (i.e., largely behavioral consultation) has evolved to encompass a wider range of topics and emphases. These include relationship development (Sheridan et al. 1992); social influence (Erchul et al. 2014); functional behavior assessment (FBA), functional analysis and brief experimental analysis (Martens et al. 2014); evidence-based interventions (EBIs; Kratochwill and Stoiber 2002); intervention plan integrity (IPI; Noell and Gansle 2014) or treatment integrity (Sanetti and Fallon 2011); multicultural issues (Ramirez et al. 1998); and larger, systems-level emphases (Sheridan and Kratochwill 2008). Given this evolution, rather than present a detailed review of traditional behavioral consultation, next a look at the process of contemporary problem-solving consultation in schools is presented (e.g., Erchul and Young 2014; Kratochwill et al. 2014). This description builds on the work of many others, including Bergan and Kratochwill (1990), Erchul and Martens (2010), Gutkin and Curtis (2009), and Zins and Erchul (2002).

The Process of Problem-Solving Consultation

Consultation incorporates a sequence of steps consisting of: relationship development; identification/clarification and analysis of the problem; intervention selection and implementation; assessment of intervention effectiveness; and planning for generalization, maintenance, and follow-up (Bergan and Kratochwill 1990; Erchul and Martens 2010; Erchul and Young 2014). Below, each step is briefly described and the entire process is summarized in Table 1. Although the key tasks found within each step are outlined in terms of a consultant/consultee dyad, they are also relevant to the work of problem-solving teams (PSTs) commonly used to implement RTI.

Establishing a Cooperative Partnership Relationship building is a critical element of problem-solving consultation (Kratochwill et al. 2014). When a consultant and consultee meet initially, each tries to become better acquainted with the other and together they attempt to develop a climate of trust and shared respect. It is important to establish, display, and maintain an awareness of multicultural issues at this stage (Ingraham 2014). Also important is to put in place an interview framework in which a consultee feels comfortable enough to initiate and respond to issues of mutual concern (Erchul and Martens 2010). Central to developing this partnership is for the consultant to avoid the *egalitarian virus*. Barone (1995) generated this term after reflecting on his failure to confront a teacher's ill-informed statement that "peanut butter cures attention-deficit

Table 1 Essential tasks within the consultation process. (Freely adapted from Bergan and Kratochwill (1990), Erchul and Martens (2010), Harvey and Struzziero (2008), and Zins and Erchul (2002))

Establishment of cooperative partnerships
Develop facilitative, respectful relationships with consultees
Establish a relational framework in which consultees are free to respond to and elaborate on issues of mutual concern
Attend to relevant multicultural considerations throughout the consultative process
Promote an understanding of and obtain agreement on each participant's roles and responsibilities
Avoid the egalitarian virus
Problem identification/clarification and analysis
Define the problem(s) and goal(s) for change in measurable, operational terms and obtain agreement with consultees
Collect baseline data on problem frequency, duration, and/or intensity
Collect curriculum-based measurement data and conduct functional behavior assessment and/or brief experimental analysis as necessary
Identify antecedents that may trigger/cue the behavior
Identify consequences that may maintain the behavior
Assess other relevant environmental factors
Identify resources available to change behavior
Intervention selection and implementation
Identify evidence-based interventions (EBIs) appropriate for the identified problem(s)
Evaluate positive and negative aspects of the proposed EBIs before implementation
Select EBI(s) for implementation from alternatives identified
Clarify implementation procedures and responsibilities with consultees
Support consultees in implementing the chosen EBI(s) through demonstration, training, and ongoing feedback
Assess intervention plan integrity (IPI)
Increase IPI using performance feedback with graphed implementation rates
Intervention evaluation and follow-up
Use collected data to inform further problem-solving and decision-making
Evaluate outcomes and any unintended effects of the EBI(s)
Assess intervention transfer/generalization, maintenance, and need for continuation
Recycle and follow-up as necessary

hyperactivity disorder" (p. 35). The egalitarian virus can disable an otherwise competent consultant who mistakenly believes that: (a) the specialized expertise of one professional is always interchangeable with that of any other professional and (b) consultees' opinions always should be given full rather than due consideration.

Identifying/Clarifying the Problem It is critical to identify/clarify the problem and then have consultant and consultee agree on the resulting definition (Bergan and Kratochwill 1990). A problem needs to be defined in clear, concise, objective, and measurable terms to the extent possible so that any progress observed in solving it may be evaluated. Through a consultant's skillful questioning, paraphrasing, and summarizing, the different dimensions of a problem can be brought to a consultee's attention and then defined in measurable terms. Once a problem has been identified in this manner, meaningful goals for change can be developed. It frequently takes considerable time to complete the problem identification/clarification process because setting/contextual factors need to be understood at this stage (Zins and Erchul 2002).

Analyzing the Problem During this step, participants try to comprehend what is causing and maintaining the problem and what resources may be available to solve it. They work to generate the best hypotheses about the problem, collect baseline data, and identify antecedent and consequential behaviors that contribute to it. Within school consultation, it is increasingly important and more common to incorporate curriculum-based measurement (CBM), FBA, and/or brief experimental analysis (Martens et al. 2014). Resources that may be used to develop and implement interventions are identified at this step. These may include student assets, system supports that help a student be successful in other settings, and material and human resources (e.g., teachers, parents, peer tutors) that are available for intervention (Zins and Erchul 2002).

Selecting an Intervention and Preparing for Implementation Both consultation and RTI place a high priority on implementing EBIs with high integrity/fidelity (Erchul 2011). Over time, the accepted practice in consultation of brainstorming interventions has been supplanted by *selecting* EBIs (e.g., Dart et al. 2012). Standard sources for locating EBIs include published meta-analyses, *What Works Clearinghouse* (ies.ed.gov/ncee/wwc), *Intervention Central* (www.interventioncentral.org), and the *Collaborative for Academic, Social, and Emotional Learning* (www.casel.org). After a list of suitable EBIs is developed, participants need to assess each for its feasibility, acceptability, cost, likelihood of success, effectiveness, consequences, etc. In planning for implementation, it is advisable to specify the steps of the intervention in writing, decide how the implementation process will be checked, and develop a plan to monitor intervention plan/treatment integrity (Erchul and Young 2014).

Implementing the Intervention The selected intervention(s) now can be implemented according to proposed plans and timelines. Consultants should be accessible to consultees to help when unforeseen problems emerge or when consultees require professional support more generally. IPI monitoring needs to occur during this step, which can be accomplished through direct observation, self-report, and/or permanent product methods. IPI has been found to increase when consultees receive specific feedback on their intervention implementation performance, and the consistency of this effect is enhanced when the data are presented in the form of graphs (Noell and Gansle 2014).

Evaluating Intervention Effectiveness and Follow-Up There are several tasks found in this stage, including determining intervention effectiveness, enhancing transfer/generalization, fading, and following up. In looking at intervention effectiveness, it is nearly always the case that the same data collection procedures used to gauge baseline performance will be used again as a basis for assessing outcomes. Evaluation methods (e.g., single-subject designs) may indicate that (a) the intervention resulted in the original goals being achieved, thereby indicating a need for follow-up monitoring, generalization, and/or fading or (b) the desired outcomes were not reached entirely, suggesting the need to recycle through earlier steps of the problem-solving process (Zins and Erchul 2002).

Team-Based Problem-Solving Consultation

Because much of the practice of RTI/MTSS is carried out in a team-based format rather than in a consultant/consultee dyad (Erchul 2011), aspects of team-based problem-solving consultation are briefly reviewed. In the 1980s, prereferral intervention teams (PITs) were formally introduced and became known as the first institutionalized team-based consultation service delivery models (e.g., Graden et al. 1985). PITs were designed to help teachers develop and implement interventions and document their effectiveness in general education. PITs were originally thought of as a means to deliver important resources to struggling children before their possible referral for special education services (Bennett et al. 2012). PITs—or similar group structures—are recommended or required by most states (Truscott et al. 2005).

Although team-based consultation approaches differ in terms of composition, model of consultation employed, and even the name used to label the teams, they generally are based in consultation theory and herald prevention as a core goal (Erchul and Young 2014). Through the years, as special education law has evolved and new regulations have been adopted, PSTs have emerged as the primary means to implement RTI. Although founded on the same basic consultation principles as PITs, PSTs usually adopt a behavioral consultation perspective and emphasize data-based decision-making (Burns et al. 2008).

PITs and PSTs have similar group membership, purposes, and activities (Erchul and Martens 2010). Both types of teams often consist of general and special education teachers, school counselors, social workers, specialists (e.g., speech, OT, PT), administrators, parents, and school psychologists. The multidisciplinary structure allows members to bring unique training experiences and professional perspectives to bear on the problems presented. To be effective, teams should be trained in consultation principles (i.e., indirect service delivery, collaboration, ecological/systems perspective) as well as the importance of adhering to a specific problem-solving model (Fuchs and Fuchs 1989). Following EBI selection/development, teachers (with assistance from team members) are usually the participants deemed responsible for intervention implementation. Teams reconvene to evaluate the effectiveness of interventions after a reasonable period of time (Bennett et al. 2012; Erchul and Young 2014).

Response to Intervention

RTI can often be seen as a school transformation model or a model of installing the systems and infrastructure to effectively meet the academic and behavioral needs of all students. The installment of the systems and infrastructure is usually led and engaged in by an implementation team, often referred to as the RTI team. Several critical components have been identified for effective RTI implementation (Batsche et al. 2005; Reschly and Bergstrom 2009; Tilly 2008; VanDerHeyden et al. 2007). The first of these is the implementation of evidence-based instruction and interventions that are increasingly intensive within a tiered system. Although different models exist in terms of the number of tiers, for the purpose of this chapter, a three-tiered model of instruction is described. Core instruction at tier 1 utilizes evidence-based practices that are implemented for all students. The goal of the core instruction is to prevent problems and to effectively meet the majority of student academic and behavioral needs. For those students identified needing additional instruction and support, increasingly intensive instruction is provided through EBIs implemented with high integrity within the secondary and tertiary tiers. Instruction is intensified in various ways including, but not limited to, the variables of frequency, duration, specialist level of the instructor, etc. Individual or small group goals are set for the intensified instruction based on national, state, and/or local standards or benchmarks.

Problem-solving is often used synonymously with RTI; however, it also may be viewed as a critical component of the RTI process (Reschly and Bergstrom 2009). Problem-solving within an RTI framework refers to a data-driven strategy encompassing the four stages of problem identification, problem analysis, plan implementation, and plan evaluation. An assumption of the problem-solving process is that no given intervention will be effective for all students (Tilly et al. 1999). Reschly and Bergstrom (2009) have described problem-solving as "pragmatic, data-based, atheoretical, and self-correcting" (p. 435). Changes are made in instructional practices and interventions based on monitoring the collected data; thus, the problem-solving process is sensitive to individual student differences. The process is utilized and increasingly intensive within the tiers of the RTI framework, such as by who is engaging in the problem-solving process (e.g., general education teacher and parent versus team of multidisciplinary instructional staff and specialists), time, etc.

It should be noted that several models of RTI do not utilize a problem-solving process as described here but instead use a standard protocol approach (Gresham 2007). In the standard protocol approach, the same empirically validated intervention is used for all students with similar academic or behavioral needs based on data. Although not tailored for individual needs, it does facilitate quality control. RTI models in practice often include the component of problem-solving (Reschly and Bergstrom 2009; Shapiro 2009). Thus, for the purpose of this chapter, when referencing RTI it is assumed that problem-solving is involved, thus making concepts from the consultation literature highly relevant.

A third identified critical component of an RTI framework is the use of instructionally relevant assessment for three purposes: screening, diagnosing, and progress monitoring. Universal screening is commonly defined as a systematic process of assessing the most basic and predictive academic skills for all students to identify students at risk for poor learning outcomes or challenging behaviors. Universal screening tests are reliable, valid, typically brief, and normed (Fuchs et al. 2007; Jenkins et al. 2007). Universal screening typically takes place three to four times in an academic year. In addition to identifying students who may be at risk, universal screening data are often utilized in evaluating the effectiveness of the core instructional program being delivered in addition with other data sources. Diagnostic assessments are utilized throughout all tiers of the model to pinpoint areas of difficulties or gaps in students' learning such that interventions can be targeted specifically to the area of need (Nelson et al. 2003). In addition, diagnostic assessments are utilized to confirm or refute hypotheses generated during the problem analysis stage of the problem-solving process. Finally, progress monitoring is the research-based practice of regularly assessing students' academic performance with brief measures used to evaluate students' response to instruction/interventions (McMaster and Wagner 2007). Progress monitoring includes setting an identified SMART (i.e., specific, measurable, attainable, realistic, and time-bound) goal, identifying the expected and actual rate of improvement, and visually representing (e.g., graphing) observed progress. Ideally, progress monitoring assessments meet the minimal standards of being reliable, valid, brief, and sensitive to growth (Fuchs and Fuchs 2008). The results of these assessments and others within the RTI model are utilized for data-based decision-making engaged in throughout the problem-solving process within all of the tiers.

In addition to these critical components, several others have been identified in practice for the successful implementation of RTI. These include leadership, professional development, and parental involvement. From the implementation science literature, it is clear that successful RTI implementation requires the application of different types of leadership to be utilized at various stages of implementation. For example, transformational leadership is critical while a school is in the exploration stage of implementation and is building support, consensus, and buy-in for implementation of RTI (Fixsen et al. 2005). Professional development with coaching is also necessary for the attainment and sustainability of RTI skills and practices (Fixsen et al. 2005; Vaughn and Fuchs 2003). Research has shown that skill training (even incorporating skill practice and feedback in training) is not sufficient within the classroom, and that ongoing support through coaching and consultation is needed for actually using the new skill in the classroom (Joyce and Showers 2002).

Finally, parental involvement is a critical component of RTI. Despite research showing strong correlations between academic achievement and parental involvement, parents are commonly involved currently in RTI initiatives through superficial and obligatory means such as parental notification (Burns and Gibbons 2008; Carlson and Christenson 2005; Lau et al. 2005). Various models exist for how to best include parents in the RTI process. These include Christenson and Sheridan's (2001) model of family–school RTI in which the intensity of the collaboration and problem-solving activities between parents and educators increases throughout the tiers. An integral component of this model is the establishment of the setting conditions necessary for effective collaboration efforts between schools and families within the tier 1 framework. These setting conditions include variables such as climate, shared goals, expectations for involvement, and recognition and value of the culture values for learning, and recognition and value of the family–school relationship (Christenson and Sheridan 2001).

In summary, RTI consists of several critical components that include: evidence-based instructional and intervention practices that are increasingly intensive; instructionally relevant assessments for the purposes of universal screening, diagnosing, and progress monitoring; data-based decision-making; leadership; professional development; and parental involvement. Effective RTI implementation is dependent on the evidence-based practices and skills called for by the critical components to be implemented as intended or with fidelity. As noted, this construct is also referred to as treatment integrity (Sanetti and Fallon 2011) or intervention plan integrity (IPI; Noell and Gansle 2014). Regardless of terminology used, however, it is necessary to achieve treatment integrity within RTI to improve student achievement and other outcomes (Erchul 2013; Reschly and Bergstrom 2009).

Effective implementation of the various RTI components has been shown to lead to several positive system-level outcomes, including improved student achievement, reduction in number of students needing intensive instruction, and reduction in number of special education referrals (Burns and Gibbons 2008; Shapiro and Clemens 2009). Effects have been found for specific subgroups of students as well. For example, VanDerHeyden et al. (2007) reported that African American students respond more quickly than other ethnic groups. Batsche et al. (2006) reported a significant decrease in the risk indices for English language learner (ELL) and African American students. Hughes and Dexter's (2011) review of 16 field studies on RTI revealed overall increased student achievement attributed to RTI implementation, with greater support for improvement in reading than mathematics outcomes; stability in reduction of special education referrals and placement rates, however, has yet to be determined. The research literature is expansive in terms of supplying support for different components of RTI, such as problem-solving and utilization of EBIs, but research is still needed mainly for examining the integration of the components into a unified framework (Burns et al. 2005; VanDerHeyden et al. 2007).

Given the impact of RTI implementation on school effectiveness, national implementation rates continue to increase (GlobalScholar 2012). Since 2008, GlobalScholar (formerly Spectrum K12) has administered a survey nationally to districts to assess the current status of RTI implementation. Results from the 2012 administration revealed that the majority of respondents (94 %) are in some stage of RTI implementation. Full RTI implementation has steadily increased from less than 10 % of respondents in 2008 to 49 % of respondents in 2012. Full implementation is defined as all of the essential components of RTI in place across all schools in core academic areas and behavior. Increases continue to be steadily seen in the percentage of districts reporting to be piloting RTI implementation as well (15 % in

2011 to 22% in 2012). Other key trends from survey results include: (a) implementation at the elementary level continues to be more widespread than at the secondary level, although implementation at the secondary level also has increased in recent years (up 28% since 2011); (b) focus of implementation continues to be largely in reading, but mathematics and behavior are increasing as well; (c) the percentage of districts reporting that general education is leading RTI implementation efforts has steadily increased; and (d) the percentage of districts with sufficient data reported increases in student achievement outcome data and reduction in the number of referrals to special education.

Relationship of Problem-Solving Consultation to RTI/MTSS

Problem-solving consultation and RTI share several major elements, yet the professional literature within each of these service delivery models rarely has informed the other (Erchul 2011). Areas of overlap/similarity in school consultation and RTI include the importance placed on: (a) the process of problem-solving; (b) prevention, broadly defined; (c) scientific, evidence-based practices; and (d) implementation of interventions with high levels of integrity.

However, there are clear points of contrast between the literatures of RTI and school consultation (Erchul 2011). Some of these differences include: (a) client focus, with RTI considering primarily the student and consultation shifting its focus between system, classroom, consultee, and student; (b) internality/externality of consultant, with RTI decidedly internal and consultation presenting a history of external change agents; (c) level at which service is delivered, with RTI commonly played out through teams and consultation through consultant–consultee dyads; (d) terminology used in each literature, with RTI tending toward more specific language; (e) confidentiality, with RTI having little to no expectation of confidentiality of communication between/among team members; (f) systematic data collection, with RTI consistently embracing this activity and consultation less so; and (g) teacher choice to participate, with RTI offering little to no choice because teachers must implement EBIs with high treatment integrity and they cannot reject EBIs if perceived as unfamiliar, time consuming, or philosophically unappealing (Martens et al. 2014).

Acknowledging these differences, increased cross talk between those who contribute to the RTI and school consultation literatures will no doubt result in tangible gains for both sides (Erchul 2011). As just one example drawn from the preceding similarities/differences analysis, the field of RTI would benefit from a greater consideration of consultee-centered issues (e.g., relationship development, teacher support, and buy-in), and the field of consultation would benefit from a greater consideration of client-centered issues (e.g., individual student assessment and outcomes). Another way that the RTI and consultation literatures may be bridged is by continuing to explore the connections between RTI and established models of consultation, such as behavioral (Bergan and Kratochwill 1990) and instructional (Rosenfield et al. 2014) consultation. These consultation models are often implemented within the RTI framework via PSTs, and because RTI is largely a team-based approach, the group consultation and PIT literature could potentially help RTI stakeholders improve the RTI process by placing importance on best practices in group dynamics or group functioning (Erchul and Young 2014).

Areas for Future Research

Despite a relatively large research base for school-based problem-solving consultation at an individual level and an emerging one at a group/team level, there are many research topics and questions that remain unexplored in this new RTI/MTSS era. Along these lines, recent randomized clinical trials of consultation (e.g., Fabiano et al. 2010; Sheridan et al. 2012) provide encouragement for future empirical efforts. Next, a sampling of possible research areas that consider the intersection of problem solving, consultation, and RTI is offered.

The first area of needed research is more systematic investigation of the use of problem-solving consultation within the larger context of RTI. Despite the numerous investigations on the essential components of RTI and complexity of such research, further research is needed to examine the integration of the components into a single, unified framework (Burns et al. 2005; VanDer-Heyden et al. 2007). This type of research would allow for systematic investigation of the essential components and the differential impact of their effects on diverse types of outcomes. Also, the examination of problem-solving consultation at an individual level or a team level integrated with other components of the RTI framework is needed. Questions that may be answered by this type of research may be along the lines of, does utilizing problem-solving consultation practices within the various RTI processes result in greater effects?

Relatedly, the RTI framework and literature do not place as much emphasis on building relationships as a means to facilitate change in teacher behavior. Instead, most commonly discussed and evaluated has been the attainment of consensus among staff and leadership on the core principles of RTI. The implementation science literature, as well as other change models of implementation literature, have emphasized the consensus gaining process as a critical component of implementation often engaged in the within the initial stages of implementation (Batsche et al. 2007; Castillo et al. 2010; Fixsen et al. 2005). Although critical in the initial stages, consensus building often does not stop in practice due to staff turnover, changes in leadership, etc. Evaluation of the impact of the relationships formed as framed in consultation during the implementation of RTI processes and its contribution to sustainability of implementation is needed. For example, does greater emphasis on the establishment of relationships and cooperative partnerships utilized

Table 2 Problem-solving consultation and RTI/MTSS: Some implications for practice

Incorporation and utilization of problem-solving consultation practices within RTI/MTSS
Professional development on the evidence-based practices of problem-solving consultation, such as indirect service delivery, collaboration, and how it differs from coaching and other related roles
Greater incorporation of practices involving cooperative partnerships, including professional development on topics such as effective group facilitation and social influence
Inclusion of teacher acceptance as a variable to be considered when problem-solving teams are at the stage of selecting EBIs, as increased teacher acceptance could lead to greater implementation fidelity of instructional strategies and interventions
Greater attention given to and emphasis placed on culturally proficient instructional strategies and multicultural issues within all tiers of the RTI/MTSS framework
Utilization of follow-up monitoring and/or generalization and fading procedures, as greater inclusion of these procedures could help facilitate more fluid movement between the varying levels/tiers of support. In addition, it may address comments often heard in practice from various stakeholders such as, "He has been in RTI forever"
Greater emphasis placed on monitoring intervention plan integrity (IPI); due to its effectiveness, RTI/MTSS teams should deliver feedback in graph form about IPI to instructors and interventionists
Establishment of clearly defined communication protocols, including parameters around confidentiality between and among RTI/MTSS team members as well as stakeholders
Increased involvement of parents within and throughout the RTI/MTSS framework to move beyond surface-level participation. To effectively do so, RTI/MTSS teams should be provided with professional development, resources, and guidelines for doing so such as those provided in Christenson and Sheridan's (2001) family–school RTI model and/or Sheridan and Kratochwill's (2008) conjoint behavioral consultation
Greater utilization of RTI/MTSS practices within problem-solving consultation
Utilization of systematic data-based decision-making and rules for guiding instructional decisions
Examination of systemic issues such as such as the effectiveness of core instruction for all students or subgroups of students for a greater preventive focus. These larger systemic issues and their impact need to be taken into account when engaging in the consultation process to develop strategies and select interventions to address needs

RTI response to intervention, *MTSS* multi-tiered systems of support within problem-solving consultation (in addition to skill attainment) result in greater likelihood of changes in teacher behavior than building consensus on principles and skill attainment?

As stated by Noell and Gansle (2014), successful implementation of the critical components of RTI (as well as various components of problem-solving consultation) is dependent upon treatment integrity or IPI. However, the field is still lacking greatly in feasible, reliable, and valid measures of integrity that can be easily captured on an ongoing basis. It is unclear from the literature which elements or constructs of fidelity or treatment integrity are critical to obtaining desired outcomes. For example, does the fidelity construct of dosage account for more of the variance in treatment outcome than the construct of teacher acceptance? If so, one could conclude that teacher acceptance of the instructional practice does not carry as much "weight" or is as critical as the dosage of the instructional practice received. Research is needed to more clearly delineate the relationships between these constructs and their effects on important outcomes of RTI/MTSS.

Another fruitful area of research is how to best train school personnel for carrying out their roles within RTI/MTSS. On the interventionist side, it is acknowledged that IPI can be enhanced through direct instruction and performance feedback (e.g., Gansle and Noell 2007) but, more generally, how can teachers be trained effectively to become competent interventionists? On the RTI team facilitator side, there is a paucity of research on how to prepare consultants in training (Newell and Newman 2014), so it is reasonable to assume that even less is known about how to effectively train individuals to guide and facilitate PSTs/RTI teams. Thus, research on training/personnel preparation for RTI/MTSS should become a clear priority for the future.

Implications for Practice

Many practice implications exist should problem-solving consultation methods continue to serve as a significant foundation for the implementation of RTI models and processes. Table 2 summarizes numerous implications along these lines.

Conclusion

The new era of RTI/MTSS requires professionals from various disciplines to step back and critically examine how research and practice of the past can meaningfully inform that of the present and future. For example, traditional consultation methods for problem-solving are an acknowledged basis for contemporary RTI/MTSS (Reschly and Bergstrom 2009). Along those lines, the primary goal of this chapter was to examine the intertwined nature of problem-solving, consultation, and RTI. Commonalities noted within the fields of consultation and RTI are the importance placed on the problem-solving process, prevention orientation, evidence-based practice, and intervention plan integrity (Erchul 2011). Still, there is much more the field of problem-solving consultation may offer to the ongoing enhancement of RTI/MTSS research and practice, and suggestions along those lines were advanced in the latter portion of the chapter.

References

Barone, S. G. (1995). The egalitarian virus. *Education Weekly, 1,* 35.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to intervention*. Alexandria: National Association of State Directors of Special Education.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2006). *Response to intervention: Policy considerations and implementation*. Alexandra: National Association of State Directors of Special Education.

Batsche, G. M., Curtis, M. J., Dorman, C., Castillo, J. M., & Porter, L. J. (2007). The Florida problem-solving/response to intervention model: Implementing a statewide initiative. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 378–395). New York: Springer.

Bennett, M. S., Erchul, W. P., Young, H. L., & Bartel, C. M. (2012). Exploring relational communication patterns in prereferral intervention teams. *Journal* *of Educational and Psychological Consultation, 22,* 187–207.

Bergan, J. R. (1977). *Behavioral consultation*. Columbus: Merrill.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy*. New York: Plenum.

Bergan, J. R., & Neumann, A. J. (1980). The identification of resources and constraints influencing plan design in consultation. *Journal of School Psychology*, *18*, 317–323.

Bergan, J. R., & Tombari, M. L. (1975). The analysis of verbal interactions occurring during consultation. *Journal of School Psychology*, *13*, 209–226.

Bergan, J. R., & Tombari, M. L. (1976). Consultant skill and efficiency and the implementation and outcomes of consultation. *Journal of School Psychology*, *14*, 3–14.

Burns, M. K., & Gibbons, K. A. (2008) *Implementing response to intervention in elementary and secondary schools*. New York: Taylor & Francis Group.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of response to intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Burns, M. K., Wiley, H. I., & Viglietta, E. (2008). Best practices in implementing effective problem-solving teams. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology—V* (pp. 1633–1644). Bethesda: National Association of School Psychologists.

Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case outcomes: Applications to research and practice. *Journal of School Psychology, 33,* 269–285.

Carlson, C., & Christenson, S. L. (2005). Evidence-based parent and family interventions in school psychology: State of scientifically based practice. *School Psychology Quarterly, 20,* 525–528.

Castillo, J. M., Batsche, G. M., Curtis, M. J., Stockslager, K., March, A., & Minch, D. (2010). *Problem solving/response to intervention evaluation tool technical assistance manual*. Tampa: University of South Florida, Florida Problem Solving/Response to Intervention Project.

Christenson, S. L., & Sheridan, S. M. (2001). *Schools and families: Creating essential connections for learning*. New York: Guilford.

Dart, E. H., Cook, C. R., Collins, T. A., Gresham, F. M., & Chenier, J. S. (2012). Test driving interventions to increase treatment integrity and student outcomes. *School Psychology Review, 41,* 467–481.

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology, 78,* 107–126.

Erchul, W. P. (2011). School consultation and response to intervention: A tale of two literatures. *Journal of Educational and Psychological Consultation, 21,* 191–208.

Erchul, W. P. (2013). Treatment integrity enhancement conceptualized as an exercise in social influence. *Journal of Educational and Psychological Consultation, 23,* 300–306.

Erchul, W. P., & Martens, B. K. (2010). *School consultation: Conceptual and empirical bases of practice* (3rd ed.). New York: Springer.

Erchul, W. P., & Schulte, A. C. (2009). Behavioral consultation. In A. Akin-Little, S. G. Little, M. Bray, & T. Kehle (Eds.), *Behavioral interventions in schools: Evidence-based positive strategies* (pp. 13–25). Washington, DC: American Psychological Association.

Erchul, W. P., & Sheridan, S. M. (2014). Overview: The state of scientific research in school consultation. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 3–17). New York: Routledge/Taylor & Francis Group.

Erchul, W. P., & Young, H. L. (2014). Best practices in school consultation. In A. Thomas & P. L. Harrison (Eds.), *Best practices in school psychology—6* (pp. 449–460). Bethesda: National Association of School Psychologists.

Erchul, W. P., Grissom, P. F., Getty, K. C., & Bennett, M. S. (2014). Researching interpersonal influence within school consultation: Social power base and relational communication perspectives. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 349–385). New York: Routledge/Taylor & Francis Group.

Fabiano, G. A., Vujnovic, R. K., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Pariseau, M. E., Naylor, J., Robins, M. L., Carnefix, T., Greiner, A. R., & Volker, M. (2010). Enhancing the effectiveness of special education programming for children with attention deficit hyperactivity disorder using a daily report card. *School Psychology Review, 39,* 219–239.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Fuchs, D., & Fuchs, L. S. (1989). Exploring effective and efficient prereferral interventions: A component analysis of behavioral consultation. *School Psychology Review, 18,* 260–283.

Fuchs, L. & Fuchs, D. (2008). The role of assessment within the RTI framework. In D. Fuchs, L. Fuchs, & S. Vaughn (Eds.), *Response to intervention: A framework for reading educators* (pp. 27–49). Newark: International Reading Association.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73,* 311–330.

Gansle, K. A., & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 244–251). New York: Springer.

GlobalScholar, Inc. (2012). *Response to intervention adoption survey*. Bellevue: GlobalScholar, Inc.

Graden, J. L., Casey, A., & Bonstrom, O. (1985). Implementing a prereferral intervention system: Part II. The data. *Exceptional Children*, *51*, 487–496.

Gresham, F. M. (2007). Evolution of the response-to-intervention concept: Empirical foundations and recent developments. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 10–24). New York: Springer.

Gutkin, T. B., & Curtis, M. J. (2009). School-based consultation theory and practice: The science and practice of indirect service delivery. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (4th ed., pp. 591–635). New York: Wiley.

Harvey, V. S., & Struzziero, J. A. (2008). *Professional development and supervision of school psychologists* (2nd ed.). Bethesda: National Association of School Psychologists.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior, 11*(*3*), 133–144.

Hughes, C. A., & Dexter, D. D. (2011). Response to intervention: A research-based summary. [Special issue]. *Theory into Practice*, *50*(1), 4–11.

Ingraham, C. L. (2014). Studying multicultural aspects of consultation. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., 323–348). New York: Routledge/Taylor & Francis Group.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36,* 582–600.

Joyce, B., & Showers, B. (2002). *Designing training and peer coaching: Our needs for learning*. Alexandria: ASCD

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of division 16 and the society for the study of school psychology task force. *School Psychology Quarterly, 17,* 341–389.

Kratochwill, T. R., Elliott, S. N., & Callan-Stoiber, K. (2002). Best practices in school-based problem solving consultation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology N* (pp. 583–608). Bethesda: The National Association of School Psychologists.

Kratochwill, T. R., Altschaefl, M., & Bice-Urback, B. (2014). Best practices in school-based problem-solving consultation. In A. Thomas & P. L. Harrison (Eds.), *Best practices in school psychology—6*. (pp. 461–482). Bethesda: National Association of School Psychologists.

Kratochwill, T. R., Clements, M. A., & Kalymon, K. M. (2007). Response to intervention: Conceptual and methodological issues in implementation. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 25–52). New York: Springer.

Lau, M. Y., Sieler, J. D., Muyskens, P., Canter, A., VanKeuren, B., & Marston, D. (2005) Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large Midwest urban school district. *Psychology in the Schools, 43,* 117–127.

Lewis, T. J., & Newcomer, L. L. (2002). Examining the efficacy of school-based consultation: Recommendations for improving outcomes. *Child & Family Behavior Therapy, 24,* 165–181.

Martens, B. K., DiGennaro Reed, F. D., & Magnusson, J. D. (2014). Behavioral consultation: Contemporary research and emerging challenges. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 180–209). New York: Routledge/Taylor & Francis Group.

McMaster, K. L., & Wagner, D. (2007). Monitoring response to general education instruction. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 223–233). New York: Springer.

Nelson, R. J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research and Practice, 18,* 255–267.

Newell, M., & Newman, D. (2014). Assessing the state of evidence in consultation training: A review and call to the field. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 421–449). New York: Routledge/Taylor & Francis Group.

Noell, G. H., & Gansle, K. A. (2014). Research examining the relationships between consultation procedures, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 386–408). New York: Routledge/Taylor & Francis Group.

Ramirez, S. Z., Lepage, K. M., Kratochwill, T. R., & Duffy, J. L. (1998). Multicultural issues in school-based consultation: Conceptual and research considerations. *Journal of School Psychology, 36,* 479–509.

Reddy, L. A., Barboza-Whitehead, S., Files, T. & Rubel, E. (2000). Clinical focus of consultation outcome research with children and adolescents. *Special Services in the Schools, 16,* 1–22.

Reschly, D. J., & Bergstrom, M. K. (2009). Response to intervention. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of school psychology* (4th ed., pp. 434–460). New York: Wiley.

Rosenfield, S., Gravois, T., & Silva, A. (2014). Bringing instructional consultation to scale: Research and development of IC and IC teams. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 248–275). New York: Routledge/Taylor & Francis Group.

Sanetti, L. M. H., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational & Psychological Consultation, 21,* 209–232. doi:10.1080/10474412.2011.595163.

Shapiro, E. S. (2009). *The two models of RTI: Standard protocol and problem-solving*. Bethlehem: Center for Promoting Research to Practice, Lehigh University.

Shapiro, E. S., & Clemens, N. H. (2009). A conceptual model for evaluating systems effects of RTI. *Assessment for Effective Intervention, 35,* 3–16.

Sheridan, S. M., & Kratochwill, T. R. (2008). *Conjoint behavioral consultation: Promoting family-school connections and intervention* (2nd ed.). New York: Springer.

Sheridan, S. M., Salmon, D., Kratochwill, T. R., & Rotto, P. J. C. (1992). A conceptual model for the expansion of behavioral consultation training. *Journal of Educational and Psychological Consultation, 3,* 193–218.

Sheridan, S. M., Welch, M., & Orme, S. F. (1996). Is consultation effective? A review of outcome research. *Remedial and Special Education, 17,* 341–354.

Sheridan, S. M., Bovaird, J. A., Glover, T. A., Garbacz, S. A., Witte, A., & Kwon, K. (2012). A randomized trial examining the effects of conjoint behavioral consultation and the mediating role of the parent-teacher relationship. *School Psychology Review, 41,* 23–46.

Sugai, G., & Horner, R. H. (2009) Defining and describing schoolwide positive behavior support. In W. Sailo, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 307–326). New York: Springer.

Tharp, R. G., & Wetzel, R. J. (1969). *Behavior modification in the natural environment*. New York: Academic.

Tilly, W. D. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-V* (pp. 17–35). Bethesda: National Association of School Psychologists.

Tilly, W. D., III, Reschly, D. J., & Grimes, J. P. (1999). Disability determination in problem solving systems: Conceptual foundations and critical components. In D. J. Reschly, W. D. Tilly, & J. P. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 285–321). Longmont: Sopris West.

Tombari, M. L., & Bergan, J. R. (1978). Consultant cues and teacher verbalizations, judgments, and expectancies concerning children's adjustment problems. *Journal of School Psychology*, *16*, 212–219.

Truscott, S. D., Cohen, C. E., Sams, D. P., Sanborn, K. J., & Frank, A. J. (2005). The current state(s) of prereferral intervention teams. *Remedial & Special Education, 26,* 130–140.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. A. (2007). Multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research and Practice, 18,* 137–146.

Zins, J. E., & Erchul, W. P. (2002). Best practices in school consultation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 625–643). Bethesda: National Association of School Psychologists.

Part II
Foundations of Practice

The Role of Professional Learning Communities in Successful Response to Intervention Implementation

Lori Helman and Kay Rosheim

Many challenges confront educational professionals as we strive to ensure that all students in our classrooms meet rigorous standards for academic achievement. Whether the goal is proficient reading by third grade, or knowledge of algebraic principles by eighth grade, or college-ready writing skills upon exit from high school, there are many obstacles that complicate the road to universal student success in school. No matter how talented or dedicated a teacher may be, the Herculean efforts of one, or even a multitude of individuals, will not be sufficient to meet important societal goals for student learning. To design instruction and school structures that help overcome impediments to achievement in specific schools, a focused and comprehensive approach within that particular context is needed. For this reason, powerful professional development is no longer seen as an isolated event for individuals, but rather a structural process in which the school becomes a learning organization with a tight focus on the success of each and every student.

We therefore judge our success in transforming the teaching profession by our students' outcomes: High levels of student achievement, judged by multiple measures that assess students' ability to understand and apply the knowledge and skills that matter most to their readiness for college, careers, and citizenship. From *Transforming the Teaching Profession Vision Statement* (US Department of Education 2012)

L. Helman (✉) · K. Rosheim
University of Minnesota, Minneapolis, USA
e-mail: lhelman@umn.edu

Professional learning communities, or PLCs, are an increasingly prevalent structure for professional development and school improvement. In PLCs, teachers, administrators, and educational specialists collaborate to understand a problem, propose and enact new ideas, and analyze the effects of their teaching on student learning. The goals of PLCs are easily aligned to a response to intervention (RTI) framework in that both are based on the tenets of teamwork, evidence-based practice, and a dedication to continuous improvement in results. This chapter takes an in-depth look at where PLCs came from, and what the research literature says about their effectiveness for improving teaching, student learning, and school productivity. Characteristics of PLCs and how these align with structures within an RTI process are shared. Next, a data team meeting that merges principles from the PLC and RTI literature is described, and examples of schools that are integrating these two currents in school reform are presented. Finally, the chapter concludes with recommendations for future research that is needed to enhance the role of PLCs in RTI implementation.

Evolution of Professional Learning Communities

In the early to mid-1990s, a transformation was taking place in traditional notions of professional development for teachers. Where once teacher development had been conceptualized

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_7

as preservice *preparation* and in-service *training,* new thinking pushed against the "pour it in" model of professional learning (Darling-Hammond and McLaughlin 1999). As student populations became increasingly diverse, teachers needed new skills, and a "packaged" approach would not be enough. Proposals for new structures and institutional arrangements were being made that included opportunities for educators to integrate theory with classroom practice, examine student work with colleagues, and engage in inquiry around problems in practice (Darling-Hammond and McLaughlin 1995). Key among the ideas being proposed was the need for procedures for reflecting critically on teaching practices and student outcomes. Darling-Hammond and McLaughlin outlined the importance of a supportive "community of practice" where teachers could closely examine and review the effectiveness of their teaching practices (1995). At the same time, Lieberman voiced similar calls for a reconceptualization of best practices for teacher professional learning (1995). She asserted that teachers "must have opportunities to discuss, think about, try out, and hone new practices…" (Lieberman 1995, p. 593). Suggestions for structuring these opportunities included common planning time, partnering new and veteran teachers, and forming teams for enacting their own professional learning (Lieberman 1995). Louis and Marks studied 24 elementary, middle, and high schools going through a restructuring process and found that a stronger level of professional community had a positive relation with the organization of classrooms for learning and students' academic achievement (Louis and Marks 1998). These ideas challenged old notions of professional improvement and set the stage for what would become the professional learning community approach.

One of the first discussions of the term *professional learning community* in print occurred in Shirley Hord's literature review entitled, *Professional Learning Communities: Communities of Continuous Inquiry and Improvement* (1997). In this document, she notes Astuto and colleagues' use of the phrase *learning community* to describe the ways in which administrators and teachers intentionally share and act upon what they are learning to improve professional practice for students' benefit (Astuto et al. 1993). In much the same way that the business sector was considering what it meant to be a learning organization (Senge 1990), a new view of school improvement was evolving: one that valued shared decision-making and collegial efforts that positioned children's learning at the center. Attributes of such PLCs included supportive and shared leadership and vision, collective learning, and the physical conditions and human resources to put these ideas into practice (Hord 1997).

DuFour and Eaker's 1998 book *Professional learning communities at work: Best Practices for Enhancing Student Achievement* outlined the rationale and purposes for PLCs. Based on DuFour's work as an educational leader at the high school level, the authors provided detailed descriptions of how these learning communities are developed and function to enhance teacher and student success. With a mixture of passion and formula, the authors heralded a new model for school change that promised to transform the culture of schools by uniting principals, teachers, and parents to work collaboratively on a shared vision for student success (DuFour and Eaker 1998).

Over the past 15 years, the literature on PLCs has expanded exponentially. A recent search of the term *professional learning community* on the Internet resulted in 242 million results that included professional organizations, publications, case studies, presentations, "how to" documents, blogs, planning tools, and more. Since Hord's literature review, many books and articles have been written on the topic, professional organizations have standardized the term, and numerous educators and consultants have made PLC implementation their full-time work. Among others, DuFour, DuFour, and Eaker disseminated a user-friendly model of PLCs across the USA (2005). The term PLC has become so ubiquitous these days that in many schools it is a synonym for collegial planning or meeting time.

Characteristics of PLCs

This section shares the essential prerequisites for productive PLCs as proposed by those who study them. What does the literature say about how they are structured for success? As the PLC model has come to life in the real work of schools, what has been learned about best practices for implementation?

In Table 1 characteristics of well-implemented PLCs based on the work of several important thinkers in the field are delineated. To a great degree, there is a consensus about the necessary components of success: shared vision, collaboration, inquiry, commitment to continuous improvement, and the appropriate physical and psychological support systems (Hord and Hirsch 2008; Kruse et al. 1994; Little et al. 2003; Stoll et al. 2006). These terms are explicated in greater depth below.

Shared Mission, Vision, Values, and Goals A PLC is unique because at its center is a collective commitment to a set of principles that govern the group's work. Foundational principles for PLCs often include putting student learning first, a willingness to share professional practice with others, and holding high expectations for all students and adults within the school community.

Collaborative Culture Professionals in a learning community work in teams that share a common purpose. Members understand that working together is more powerful than individual efforts. They learn from each other and create the momentum that drives improvement. Together, teams build the structure and tools that make collaborative work and learning effective and productive. Members of the PLC construct a set of agreed-upon norms for how the group will engage and operate.

Collective Inquiry Professionals in a learning community seek improved methods of teaching and learning. They test these methods, examine evidence of student learning, and then reflect publicly on the results. Teams coordinate their investigations and actions so that the work of each individual contributes to a common strand of new knowledge applicable in their school context (Dunne et al. 2000).

Action Focus Members of PLCs turn their learning and insights into action. They recognize the importance of engagement and experience in learning and in testing new ideas. Their collaboration moves beyond discussion about students to producing materials that improve instruction, curriculum, and assessment.

Commitment to Continuous Improvement Members of PLCs continually seek ways to bring present reality closer to future ideal. They constantly ask themselves and each other: "What is our purpose?" "What do we hope to achieve?" "What are our strategies for improv-

Table 1 Characteristics of professional learning communities

Source	Shared mission, vision, values and goals	Collaborative culture	Collective inquiry	Action focus	Commitment to Continuous Improvement (professional growth)	Physical support	Relational support (mutual trust)	Supportive leadership	Shared leadership
DuFour 2004	X	X	X	X	X	X			
Hord and Hirsh 2008	X	X	X	X	X	X	X	X	X
Kruse et al. 1994	X	X	X	X	X	X	X	X	X
Stoll et al. 2006	X	X	X	X	X	X		X	X
Little 2003	X	X	X	X	X	X	X	X	X

ing?" and, "How will we assess our efforts?" Assessment and reevaluation are the keys to continued improvement.

Physical Supports Administrators in a PLC provide the sanctioned time, space, and resources needed for PLC members to work together.

Relational Support Members of PLCs respect and value the opinions of other team members. Norms allow all members to trust their colleagues enough to voice weaknesses and questions, as well as strengths and successes. The focus is on collaboration and goal achievement, as opposed to competition.

Supportive Leadership Administrators share decision-making with teachers and provide leadership opportunities.

Shared Leadership Leadership is shared and distributed formally and informally among PLC members. All participants take responsibility for the success of the group process.

Taken together one paradigm-shifting concept emerges from the characteristics of effective PLCs: what Louis, Marks, and Kruse (1996) call the "deprivatization of practice." Deprivatization occurs when teaching is no longer seen as an activity conducted "behind closed doors" by an individual and becomes a professional practice that is studied, questioned, and replicated by a team. In PLCs, teachers share, discuss, compare, and question each other about their instructional techniques. Classroom teaching and assessment are opened up for observation and reflection, and teachers build on each other's successful practices in order to achieve school-wide goals.

Research on Professional Learning Communities

Research on the relationship of PLCs to educational outcomes focuses on three areas: (1) shifts in teaching practice, (2) student achievement, and (3) school-wide or program effects.

Teaching Practice By far, the bulk of research to date on PLCs has focused on their relationship to changes in teacher behaviors and attitudes. The literature demonstrates that teachers who participate in PLCs demonstrate more of the positive behaviors that school change projects seek to regularize. These capacities include increased leadership and collaboration skills, higher satisfaction and participation in professional activities, and improved knowledge of research-based approaches to instruction. These features are delineated over the next several paragraphs.

On a personal level, PLCs have been associated with increased teacher morale, participation, and ownership of the functioning of the school. In schools that were organized into student and teacher learning communities, Lee, Smith and Croninger (1995) found that teachers expressed greater job satisfaction and lower rates of absenteeism. In her summary of the literature based on reports of school restructuring efforts, Hord (1997) concluded that outcomes for staff working in PLCs included a reduction of feelings of isolation and an improved commitment to the mission of the school. Several researchers noted changes in teachers' attitudes after participation in collaborative professional development, such as an increased sense of confidence and efficacy, a willingness to collaborate and share (even though there was anxiousness about having their teaching observed), and an openness to trying out new practices (Cordingley et al. 2003; Supovitz 2002).

Skills for leading and collaborating also appear to be developed through participation in PLCs. PLCs facilitate a culture of collaboration that includes peer observation and feedback, collegial support, dialogue about practice, and extended time to embed new instructional strategies (Berry et al. 2005; DuFour 2004; Stoll et al. 2006). PLC members develop shared norms and values that create opportunities to build collaborative knowledge through the use of collegial dialogue to reflect upon the successes or inadequacies of their practice (Dufour 2004; Lieberman 2003; Newmann and Wehlage 1996; Phillips 2003). Vescio, Ross, and Adams' (2008) literature review found documentation that these collaborative efforts led to teachers changing their instructional practices to include more interaction (e.g., author's center, choral reading, writing process), flexible grouping strategies, and focusing on the needs of diverse students. Wilhelm (2010), reporting on a California project that implemented learning communities in close to 100 schools, noted that principals and teachers demonstrated increased levels of shared responsibility for student achievement.

Perhaps most importantly, participation in PLCs appears to engender an increased teaching focus on student learning (DuFour 2004; White and McIntosh 2007). Since essential characteristics of PLCs include collective inquiry, an action focus, and continuous improvement, team meetings are primarily centered around student performance outcomes (Clauset et al. 2008; Love et al. 2008; Peery 2011). When PLC meetings focus on student data, they are much more likely to have an impact on achievement, changes in school culture, and instructional quality (McLaughlin and Talbert 2010; Saunders et al. 2009; Vescio et al. 2008).

Student Achievement Although PLCs are frequently discussed in the research literature, few studies have measured their effect on student achievement. The Center for Comprehensive School Reform and Improvement points to four studies that attempt to measure the connection between PLCs and student performance (see http://www.centerforcsri.org/plc/literature.html). In the first study, Hughes and Kristonis (2007) analyzed data from schools in Texas that were using PLCs. Over the course of a 3-year time period, 90.3 % of schools showed an increase in mathematics scores on standardized tests, and 81.3 % of schools demonstrated an increase in English language arts scores (Hughes and Kristonis 2007). Strahan (2003) conducted case studies of three elementary schools with diverse populations that participated in PLCs and found that students from low income and ethnic minority groups improved from 50 to 75 % on proficiency levels on state achievement tests. Supovitz (2002) and Supovitz and Christman (2003) discerned that PLCs may have an effect on teachers' attitude and school culture, but without an explicit focus on instruction, there may be few gains in student achievement.

A recent quasi-experimental study by Saunders et al. (2009) involved a 5-year, two-phased investigation on the effects of grade-level team meetings for student achievement in nine experimental and six control elementary schools. In phase 1, only principals were trained in meeting protocols. In phase 2, principals and teacher-leaders were given guidance on the team meetings and explicit protocols were shared. Results showed no differences in student achievement between control and experimental groups during phase 1, but experimental groups showed faster growth and better scores on standardized assessments in phase 2 of the study. The authors conclude that shared leadership, focused teams, and clear protocols are essential to producing increased student achievement.

School Wide or Program Effects There are many articles on the positive effects that PLCs have had on overall school improvement. Observed advancements include an increased focus on data-based decision-making (Strahan 2003), a school-wide focus on results (DuFour 2004), mutual accountability of staff (Reichstetter 2006), effective scheduling of curricular blocks (Supovitz 2003), and enhancement of communication mechanisms (Burnette 2002). The website *All Things PLC* provides a forum for schools to share their successes using the PLC model (see http://www.allthingsplc.info/evidence/evidence.php). Currently, there are 170 elementary and secondary schools listed as contributing data on their school-wide practices and successes to share with educational colleagues. Unfortunately, among the criteria for posting school information to the website, schools must be able to show student achievement growth and sustainment. Thus, this site is more a showcase of anecdotal information rather than a repository of evidence on the effectiveness of PLCs.

PLCs and Response to Intervention Approaches

The next portion of this chapter connects the literature on PLCs to RTI frameworks of multi-tiered systems of support. The ways that PLCs can serve as vehicles for RTI goals, and how the data-based and collaborative decision-making processes intrinsic to RTI uphold the goal of improved student learning in team meetings, are discussed. Finally, an example structure for PLC meetings within an RTI system is described, and the types of interactions and outcomes associated with these meetings are shared.

Goals of RTI RTI involves the targeting of resources to meet all students' needs through the systematic use of assessment data and efficient allocation of resources to ensure enhanced learning (Burns and Gibbons 2008). The RTI organizational model is typically conceptualized as a three-tier support system with approximately 80% of students receiving research-based core instruction in the general education classroom (tier 1), approximately 15% of students needing additional targeted small group instruction as part of general education (tier 2), and approximately 5% of students possibly requiring intensive, focused interventions based on problem-solving models (tier 3; Burns and Gibbons 2008). The RTI model aims to prevent and remediate learning problems through the utilization of meaningful and ongoing assessment, instructional differentiation, and collaboration among knowledgeable professionals working to maximize each student's success (International Reading Association 2010).

Uniting RTI and PLCs Several guiding principles bind RTI to the work of PLCs. Both frameworks involve models for school-wide change; they acknowledge that support for improved instruction and student learning comes from whole school structures and cannot be implemented in isolation by individual teachers. Both RTI and PLC models value collaboration among multiple professional perspectives including principals, teachers, resource specialists, and school psychologists. In a similar manner to PLCs, an RTI model relies on shared values, physical and relational support, supportive leadership, and a commitment to continuous improvement. Because both the PLC and RTI structures require focused inquiry and substantial institutional support, it is critical that schools implementing these approaches integrate the processes into one cohesive school-wide plan. If not united, the strain on time, material resources, and professional energy to carry out two diverse initiatives will likely reduce the effectiveness of both efforts.

RTI is built on the need for the highest quality core instruction in the general education classroom (Vellutino et al. 1996). It is only by having an opportunity to experience research-based, well-delivered, developmentally appropriate instruction in the general education classroom that educators will know which students might require additional targeted support for academic success (Wixson et al. 2010). This core instruction should be based on the most current research in the content area, as well as teaching practices that have been found to be successful with the specific population of students in the classroom. In addition, the instruction must be based on formal and informal assessments and differentiated to students' developmental levels accordingly. Finally, for students who are learning English as a new language, classroom instruction must be tailored for accessibility by clarifying unknown vocabulary, modeling academic language structures, frequently checking students' understanding, and fostering student discussion and questioning (PRESS 2011). Effective teaching practices in core instruction are important inquiry topics for schools implementing a PLC or an RTI framework. Even better, when these two frameworks are integrated, teachers examine their instructional practices in collegial teams, discuss the effect of their teaching on student learning through examination of student work samples, and ensure that each content area or grade-level team is providing instruction that is maximizing student success.

The integration of RTI into the PLC framework gives it additional power and facilitates teacher buy-in. Once teachers have had an opportunity to inquire into their core instructional practices and work with colleagues to regularize optimal approaches, the team can evaluate whether or not such a teaching approach is meeting all students' needs. If a high percentage, but not all students are successful, the team turns its attention to what next steps can be put in place for targeted, supplementary support (tier 2) for those students who are not making good progress. The outcome of these team discussions may be to provide students with small group supplemental instruction within the classroom, a collaboration of focused instruction across a grade level or content area, or to have the general educator team with a specialist or support teacher within the school. After a short period of time, the PLC team reevaluates the progress of students receiving this tier 2 support.

Uniting RTI into a PLC framework is likely to change some of the latter's routines. Whereas some PLC teams may operate more as a study group than a data analysis team, an RTI framework will institutionalize the regular review of students' progress on key academic benchmarks. For example, at the elementary school level, in order to ensure that all students are making sufficient progress on district and state performance standards, approximately three times a year a benchmark assessment will be given to all students. Following this data collection, PLC teams will need to review the assessment results to determine which students are on track to meet academic goals, and which students need extra support to be successful. Once students have been identified to receive small-group support at the tier 2 level, monthly meetings to examine their progress must become part of the ongoing agenda of the PLC team. If a strong learning trajectory does not become apparent, the team will need to consider other tier 2 approaches or a more intense, tier 3 intervention (Burns and Gibbons 2008).

A Data-Focused Instructional PLC Meeting As outlined in the previous section, there are several ways that PLC and RTI meetings can be integrated. The integrated team meetings may involve inquiries into which teaching practices are producing the best student learning; which students are meeting or not meeting benchmark achievement goals, and what supports might be put in place for those who are struggling; or the PLC meeting may focus on assessing the progress of students who are receiving tier 2 support and making decisions about adaptations or next steps. No matter what the focus of a team meeting is, it is essential for each session to review and make connections to student learning, either by examining artifacts of student work, informal assessment data, or assessments of progress on academic benchmarks collected through regular progress monitoring (e.g., an established data collection tool such as *AIMSWeb* or *FAST*). Given the hectic nature of busy schools, a clear schedule that outlines the topics and goals of each PLC team meeting is essential to their efficiency and productivity. Regularized procedures and note-taking forms will help each member of the team stay organized and contribute to the learning of the group.

A variety of formats for team meetings have been created, many sharing very similar procedures and goals. These protocols have been created by personnel at individual schools, districts, state departments of education, PLC and RTI consultants, and institutions of higher education. Some resources for conducting data-focused PLC meetings are shared in the next section, and this section shares several typical formats, and then takes the reader into what a data-focused PLC meeting might look like in practice.

The goal of an instruction-focused data PLC meeting is to analyze student data to improve the quality of core instruction. One structure proposed for these team meetings by the *Leadership and Learning Center* involves a five-step process. After a pre-assessment has been given to students, the team meets and (1) displays the data, (2) analyzes the data and prioritizes needs, (3) sets a measurable goal and growth target, (4) determines instructional strategies to raise student proficiency, and (5) determines indicators of implementation strength (Peery 2011). The process is cyclical; after implementation, a new round of data is examined to investigate student learning. In a similar manner, the *Whole Faculty Study Group Approach* involves educators in (a) understanding the learning need, (b) collecting and analyzing baseline data, (c) establishing improvement targets, (d) selecting and implementing teacher interventions, and (e) collecting and analyzing post-intervention data (Clauset et al. 2008). The *using data project* provides intensive support to data coaches to set the groundwork for a collaborative inquiry group that drills down into data from local and state sources to identify student learning problems, causes, and potential solutions. Next, the team implements, monitors, and reviews results for achievement progress (Love et al. 2008). Key to all of these approaches are: the use of informal assessments of student learning, analyzing and setting goals, collaboratively selecting instructional best practices, and reevaluating student learning after implementation.

In the following vignette of a fifth-grade data-focused PLC meeting, the authors provide an example of how the structures identified above can be put into practice. The meeting is facilitated by a grade-level lead teacher (T1). Participants include the other three 5th grade teachers and the school reading specialist.

> The group sits down and the lead teacher reviews the group-developed norms (each meeting starts this way). Next, student work is displayed on the Smartboard and student progress is discussed child by child. On this day, the topic was a review of a formative assessment of main idea. The assessment was a short reading passage followed by five multiple-choice questions written to assess the mastery of main idea in this passage. The data are displayed by class, student, and item number.
> T1 starts the conversation, "So who wants to start? Does anyone notice anything about the data?"
> T2: "I notice that [T1]'s students as a whole did better than mine. Can you tell us what you did to prepare the students for the test? How much time have you been spending on main idea?"
> T1: "I work on main idea a lot! I think one reason why my students did well is that I just did a lesson on determining importance from our language arts curriculum. I love those concrete lessons, especially for my EL students. This is the one where I told my students I was going to basketball practice after school today, and showed them what was in my gym bag. I asked which items were important for helping me at basketball practice. I had some funny things in the bag, too, so we had fun with the lesson."
> T4: "Those are great, engaging lessons. They really seem to make sense for my students too. My question is—how did you get your students from that lesson to the paper-pencil multiple choice formative assessment that we just took? My students have trouble connecting the dots, and apparently, based on my test scores, I have trouble helping them do it."
> T1: "From my gym bag lesson, we had follow-up discussions on how the lesson could be applied to reading. Productive talk is really important. In the book there are also ideas for posters that I put up after teaching a lesson. The posters stay up the rest of the year and serve as a reference to students. We also review them periodically. I ask the students to use these in their reading response journals for both their assigned and free choice reading until I have evidence that they are understanding and mastering the concept. For determining importance, I use thinking stems such as—What's important here? I want to remember that…, One thing we should notice is… So we are talking, writing, reading, talking, writing, and reading all the time."
> T3: "Do you know if there are more copies of that book in the building?"
> Reading Specialist: "Yes, there are several copies in our professional library. I agree with you—it is a great resource. The students love the lessons because they are so engaging and fun to teach. They serve as concrete bridges to strategic reading and I think we have evidence that they have had an impact on [T1]'s students."
> T2: "Changing the subject a bit, but I noticed something in even how the answers were circled on one of my student's tests. This is the student that [reading specialist] just did more diagnostic testing on after he did so poorly on the standardized reading test. I think this student doesn't see value in these assessments and speeds through to get it done. See how the answers are circled neatly on the first 3 questions, and the last 2 are sloppily drawn? This makes me think, he rushed answering the last two questions once the other students started to hand them in. This kid thinks faster is smarter." (T2 shows the team the test).
> The reading specialist discusses her recent assessment with the student. She notes that earlier in the week she gave the student an informal reading inventory and saw the student struggle in reading multisyllabic words. He often made little attempt to decode an unknown word-saying the beginning of the word and then putting anything at the end "just to finish it." He made many careless errors, including skipping a whole line without noticing. She noticed no evidence of him reading for meaning. The reading specialist describes how she and the student chatted about the results and devised a

plan of things he would work on this week such as: checking for understanding and slowing down to make certain the words make sense. She shared her plan to do some word work with him taking apart longer words, separating base words from endings, and figuring out meanings of words from context. As the meeting time is nearly over, the team agrees to try a concrete lesson on determining the main idea and give another formative assessment to review the following week. T1 also commits to contacting the gifted support teacher to request ideas for expanding the lessons for accelerated students.

In the previous vignette, several characteristics of an instructionally focused PLC are evident. The teachers, with the support of the reading specialist, compare student work samples and discuss what each teacher did to support student learning. Teachers share their instructional practices and ask questions to each other. In addition, the artifacts of particular students who are not successful with the activity are analyzed, and the reading specialist contributes her advanced knowledge of how to support students who struggle in particular areas. Because the teachers are discussing the grade-level goals they have set—in this case understanding the main idea—and analyzing their own students' work, they are deeply invested in learning together.

A Benchmark Data PLC Meeting The RTI framework recommends universal screening of all students periodically (approximately three times per year) to make sure they are on track to meet grade level and content area goals (Burns and Gibbons 2008). After each benchmarking assessment, it is critical for school personnel to use their PLC meeting time to review the screening data and make sure that students are receiving the appropriate level of tiered support in the content area. VanDerHeyden and Burns (2010) propose a structure for the grade-level team meeting that includes the following questions:

1. Are there any class-wide problems?
2. If there are no class-wide problems, which students need a tier 2 intervention?
3. What data are needed to decide which tier 2 intervention to use with each student?
4. Is the tier 2 intervention working for each individual student receiving one?
5. Should we refer any students to the problem-solving team?
6. Is the tier 3 intervention working for each individual student receiving one?
7. Are there any students whom we should refer for a special education evaluation? (p. 130)

The benchmark data PLC meeting takes place on a regular basis following school-wide screening assessments and for regular progress monitoring of students in tier 2 and tier 3 interventions. The goal is to ensure that students are receiving a focused intervention that puts them on track to catch up to curricular benchmarks. If the intervention is not working, it is important to select and measure another instructional approach for supplemental intervention. This type of data-based PLC meeting can be efficiently handled in 20–45 min, depending on the quantity of benchmark data to be examined. It provides an excellent opportunity for classroom teachers to meet with those specialists in the school who provide intervention services to students, and to check that each student is receiving cohesive instruction in and outside of the classroom setting.

A school-wide literacy improvement project underway in a large urban district in the Midwest is one example of the use of data PLC meetings for both instructional and benchmark data purposes. The project is a partnership among a research university, an urban school district, a nonprofit organization, and a corporate sponsor. It is built on four key principles: quality core instruction, data-based decision-making, tiered interventions, and embedded professional development (PRESS 2011). All four of the core principles are realized through data-based PLC meetings. At least twice a month, teachers have opportunities to meet their PLC grade-level groups to analyze student work and informal assessments and collaboratively select instructional practices to address student needs. Once a month, the data-based PLC meets to review benchmark data for all students or progress monitoring data for students receiving tier 2 or tier 3 interventions, and make recommendation for instructional adaptations or changes in the level of intensity of the intervention. In this project, data-based decision-making is no longer in the hands of a few specialists; rather, instructional staff, specialists, and administration work together to inquire, propose, test, and validate what is working for student achievement, and how this instruction can be continuously improved.

Implications for Practice

Throughout this chapter, a number of important practices for implementing PLCs within an RTI framework are identified. Table 2 below summarizes these suggestions.

At the heart of these practical implications is the knowledge that it is up to whole faculties, along with their administrators, to work collaboratively to find ways to increase students' achievement. Teaching is no longer a "deliver-and-move-on" activity that it is hoped students internalize; rather, it involves providing quality core instruction for all, with frequent assessment and prompt assistance to catch students if they fall behind, and working relentlessly to address learning problems early (DuFour and Marzano 2011).

Resources for Implementing PLCs

A variety of print and electronic resources are available for schools to use as they implement PLCs with a focus on data-based decision-making. The following resources are not intended as an exhaustive list, but as a starting point for opening doors to the literature.

All Things PLC. http://www.allthingsplc.info/. Provides research, articles, data, and tools to educators who seek information about the *Professional Learning Communities at Work*™ process. An affiliate of Solution Tree, Inc., this information is provided so schools and districts have relevant, practical knowledge and tools as they create and sustain their PLCs.

The Center for Comprehensive School Reform and Improvement. http://www.centerforcsri.org/plc/. Provides an extensive bibliography and links to background information, research,

Table 2 Implementing PLCs within an RTI framework

Implication for practice	Explanation
Engaging in data-based PLC meetings requires that a set of prerequisites are in place at the school including shared vision, collaboration, inquiry, commitment to continuous improvement, and the appropriate physical and psychological support systems	Without these foundational characteristics, PLCs have not been found to be critical to school improvement
The work of the PLC must be based on improvements in teaching and learning that are measurable. Artifacts of instruction and student work need to be a part of each collegial meeting	The products of teaching and learning take group discussions from the realm of the theoretical to the possibility of replication in multiple settings
Integrating RTI and PLC frameworks allow schools to focus their efforts for maximum professional learning and instructional effect	Time and resources in schools are limited and valuable; initiatives must be focused and efficient
A schedule for data-based PLC meetings should consist of at least one opportunity a month to discuss benchmarking and progress monitoring data to review the achievement of students who need extra support to meet grade-level standards	Students need early and targeted support before they fall too far behind
Data-based PLC meetings should also include one or more opportunities a month for staff to examine student learning, and select and evaluate promising teaching practices to use in class. A cycle is developed for each new practice to be evaluated by the team in an upcoming PLC meeting	Professional development is most powerful when it is connected to student learning and brings together the expertise of a group of professionals

PLC professional learning community, *RTI* response to intervention necessary supports, and numerous articles on implementation of PLCs.

Corwin Publishing Company. http://www.corwin.com/topics/C89. One of the leading publishers on the topic of PLCs, they also offer webinars related to the topic.

Educational Testing Service: Keeping Learning on Track. www.ets.org/kit. *Keeping Learning on Track* is a sustained, interactive professional development program that helps teachers adopt minute-to-minute and day-by-day assessment-for-learning strategies. It is the result of a 3-year research and development process led by the author and ETS's *Learning and Teaching Research Center*.

Learning Forward. http://www.learningforward.org/standards/learning-communities#.UMVbVoVD-Hk. Formerly the National Staff Development Council, this professional development organization outlines standards for learning communities, resources, webinars, and guides to implementation.

Pearson Learning Teams: Professional Learning Communities Guided for Results. http://www.pearsonlt.com/about/implementation. Provides an overview and resources related to PLCs—their five core elements, research base, successes, and how they are implemented.

SEDL: Advancing Research, Improving Education. http://www.sedl.org/pubs/change34/. Provides links to research articles, books, professional development, surveys, and case studies on the work of PLCs for continuous school improvement.

Areas for Future Research

This chapter describes the work of PLCs and their connection to an RTI framework. The use of PLCs in staff development and school improvement is prolific, although many questions about how to ensure that PLCs are productively used to advance student learning remain unanswered. This section highlights areas for future research that will fortify the implementation of data-based PLC meetings.

To begin, there are significant variations in terminology as to what constitutes a PLC. It is imperative that defining characteristics of PLCs are clearly explicated so that these criteria can be replicated to increase the possibility of success in other sites. A consensus from major investigators, followed up with documentation and testing in the field, will be critical to move practice forward.

Next, as previously noted, there is little evidence that working in PLCs has an impact on student achievement. Most data on PLCs involve measurement of changes in teaching beliefs and practices, and these are typically self-report data. The few studies noted in this chapter that examine student outcomes must be augmented by well-conceived and methodologically-rigorous investigations in schools. Saunders et al.'s (2009) quasi-experimental study using experimental and control schools provides an excellent example of the kind of research that is needed before scaling up takes place. Much of the research on the benefits of PLCs took place after schools were identified as successful and school improvement was progressing. Future research should look to schools that are yet to implement these procedures, so that baseline data are available and can be compared to nonparticipating schools.

Another important area for future research involves the information that practitioners—teacher leaders, principals, coaches, and others—need to know in order to successfully implement data-based PLCs. Some of the questions that research might address include: What is the most effective role for a content specialist as a member of a PLC? What strategies are best introduced when the team does not have the expertise to address a challenge in student learning that it confronts? How can teachers be supported to provide the differentiation that students need? What types of data are the most important to be collected and analyzed at data-based PLC meetings? These and similar questions are the nuts and bolts of successful teamwork and student-focused action.

Finally, another critical research area involves developing team expertise in issues of cultural and linguistic variation relating to student achievement. Professional capacity to understand difference, as opposed to deficiency, is a key component for PLCs (Lindsey et al. 2009). The field will profit greatly from those who can situate learning about teaching and student achievement within a multicultural and multilingual perspective.

Conclusion

This chapter examines the evolution of PLCs and the research literature on their effectiveness for improving teaching, student learning, and school productivity. Characteristics of PLCs are outlined and how they fit into the RTI structure for addressing all students' needs is illustrated. A data team meeting that merges principles from the PLC and RTI literature is provided, and resources to call upon for implementing data-based PLCs are shared. In many settings, PLCs have created synergy among staff and school leadership to tackle pressing school-wide instructional and achievement issues. Building on what one knows to date, collaborative teams that are focused on data analysis and action research may be a powerful force to support increased student learning.

References

Astuto, T. A., Clark, D. L., Read, A.-M., McGree, K., & Fernandez, L. deK. P. (1993). *Challenges to dominant assumptions controlling educational reform*. Andover. Regional Laboratory for the Educational Improvement of the Northeast and Islands.

Berry, B., Johnson, D., & Montgomery, D. (2005). The power of teacher leadership. *Educational Leadership, 62*(5), 56–60.

Burnette, B. (2002). How we formed our community. *Journal of Staff Development, 23*(1), 51–54.

Burns, M. K., & Gibbons, K. A. (2008). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. New York: Routledge.

Clauset, K. H., Lick, D. W., & Murphy, C. U. (2008). *Schoolwide action research for Professional Learning Communities*. Thousand Oaks: Corwin Press.

Cordingley, P., Bell, M., Rundell, B., & Evans, D. (2003). The impact of collaborative CPD on classroom teaching and learning. *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*(8), 597–604.

Darling-Hammond, L., & McLaughlin, M. W. (1999). Investing in teaching as a learning profession: Policy problems and prospects. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 376–411). San Francisco: Jossey-Bass.

DuFour, R. (2004). What is a Professional Learning Community? *Educational Leadership, 61*(8), 6–11.

DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington: Solution Tree Press.

DuFour, R., & Marzano, R. J. (2011). *Leaders of learning: How district, school and classroom leaders improve student achievement*. Bloomington: Solution Tree.

DuFour, R., DuFour, R., & Eaker, R. (2005). *On common ground: The power of professional learning communities*. Bloomington: Solution-Tree.

Dunne, F., Nave, B., & Lewis, A. (2000). Critical friends groups: Teachers helping to improve student learning. *Phi Delta Kappa International Research Bulletin (CEDR), 28,* 9–12.

Hord, S. (1997). Professional learning communities: Communities of continuous inquiry and improvement. http://www.sedl.org/pubs/change34/. Accessed 3 Dec 2012.

Hord, S. M., & Hirsch, S. A. (2008). Making the promise a reality. In A. M. Blankstein, P. D. Houston, & R. W. Cole (Eds.), *Soul of educational leadership: Sustaining professional learning communities* (pp. 23–40). Thousand Oaks: Corwin Press.

Hughes, T. A., & Kritsonis, W. (2007). Professional learning communities and the positive effects on student achievement. A national agenda for school improvement. http://www.allthingsplc.info/pdf/articles/plcandthepositiveeffects.pdf. Accessed 8 Dec 2012.

International Reading Association. (2010). *Response to intervention: Guiding principles for educators from the International Reading Association* (Position statement). Newark: Author. http://www.reading.org/general/AboutIRA/PositionStatements/ResponsetoIntervention.aspx. Accessed 10 Dec 2012.

Kruse, S., Louis, K. S., & Bryk, A. (1994). Building professional community in schools. *Issues in Restructuring Schools,* (6), 3–6.

Lee, V. E., Smith, J. B., & Croninger, R. G. (1995). Another look at high school restructuring. *Issues in Restructuring Schools, Fall* (9), 1–10.

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. *Phi Delta Kappan, 76*(8), 591–596.

Lieberman, A. (2003, August). Building learning communities in underperforming schools: The tension between improvement and survival: Seattle case study summary. www.teachercollaboration.net. Accessed 10 Dec 2012.

Lindsey, D. B., Jugwirth, L. D., Pahl, J. V. N. C., & Lindsey, R. B. (2009). *Culturally proficient learning communities*. Thousand Oaks: Corwin Press.
Little, J. W., Gearhart, M., Curry, M., & Kafka, J. (2003). Looking at student work for teacher learning, teacher community, and school reform. *Phi Delta Kappan, 85*(3), 184–192.
Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33*(4), 757–798.
Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom teacher's work and student experiences in restructuring schools? *American Journal of Education, 106*(4), 532–575.
Love, N., Stiles, K. E., Mundry, S., & DiRanna, K. (2008). *The data coach's guide to improving learning for all students*. Thousand Oaks: Corwin Press.
McLaughlin, M. W., & Talbert, J. E. (2010, Spring). Professional learning communities: Building blocks for school culture and student learning. Annenberg Institute for School Reform VUE. http://www.annenberginstitute.org/VUE/. Accessed 10 Dec 2012.
Newmann, F. M., & Wehlage, G. G. (1996). Conclusion: Restructuring for authentic student achievement. In F. M. Newmann (Ed.), *Authentic achievement: Restructuring schools for intellectual quality* (pp. 286–301). San Francisco: Jossey-Bass Publishers.
Peery, A. (2011). *The data teams experience: A guide for effective meetings*. Englewood: Lead + Learn Press.
Phillips, J. (2003, Spring). Powerful learning: Creating learning communities in urban school reform. *Journal of Curriculum and Supervision, 18*(3), 240–258.
PRESS. (2011). Path to Reading Excellence in School Sites (PRESS) Resources. http://www.cehd.umn.edu/reading/PRESS/resources/default.html. Accessed 6 Dec 2012.
Reichstetter, R. (2006). *Defining a professional learning communitiy. E & R Research Alert,* (No. 06.05).
Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing achievement by focusing grade level teams on improving classroom learning: A prospective, quasi-experimental study of Title 1 schools. *American Educational Research Journal, 46*(4), 1006–1033.
Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*(4), 1–38.
Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal, 104*(2), 127–146.
Supovitz, J. A. (2002, December). Developing communities of instructional practice. *Teachers College Record, 104*(8), 1591–1626.
Supovitz, J. A., & Christman, J. B. (2003). Developing communities of instructional practice: Lessons for Cincinnati and Philadelphia. CPRE Policy Briefs, University of Pennsylvania, 1–9.
U.S. Department of Education. (2012). Transforming the teaching profession. www2.ed.gov/documents/labor-management-collaboration/vision-statement.pdf. Accessed 24 Oct 2012.
VanDerHeyden, A. M., & Burns, M. K. (2010). *Essentials of response to intervention*. Hoboken: Wiley.
Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S., Chen, R., Pratt, A., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.
Vescio, V., Ross, D., & Adams, A. (2008). Review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91.
White, S. H., & McIntosh, J. (2007). Data delivers a wake-up call. *Journal of Staff Development, 28*(2), 30–35.
Wilhelm, T. (2010). Fostering shared leadership. *Leadership, 40*(2), 22–38.
Wixson, K. K., Lipson, M. Y., & Johnston, P. H. (2010). Making the most of RTI. In M.Y. Lipson & K. K. Wixson (Eds.), *Successful approaches to RTI: Collaborative practices for improving K-12 literacy* (pp. 1–19). Newark: International Reading Association.

Response to Intervention and Accountability Systems

Timothy J. Runge, David J. Lillenstein and Joseph F. Kovaleski

The following are terms and concepts salient to this chapter:

Data analysis team (DAT) A group of educators and support personnel who regularly meet to review student data and make instructional and resource-allocation decisions based on those data.

Data warehousing Secure, computer-based systems of storing data that provide for efficient report generation and graphical presentations used in DAT meetings.

Report generation The capacity of secure, computer-based data warehousing systems to generate a variety of data reports that helps DAT meetings be more efficient and effective at making data-based decisions.

Visual displays The capacity of secure, computer-based data warehousing systems to generate a variety of pictorial representations of data important for review in DAT meetings. Typically, these displays are in the form of bar graphs, histograms, or time-series graphs.

T. J. Runge (✉) · J. F. Kovaleski
Indiana University of Pennsylvania, Indiana, USA
e-mail: trunge@iup.edu

D. J. Lillenstein
Derry Township (PA) School District, Hershey, USA

Data Used in RTI and PBIS

Response-to-intervention (RTI) logic is predicated on the concept that all students receive evidence-based core instruction in academic, social, and behavioral domains. The vast majority of students will develop appropriate skills when afforded this universal, core instruction often referred to as tier 1 instruction. Some students, however, do not develop these skills and require additional supports and interventions (i.e., tier 2 interventions) targeted on their skill deficits. Most students respond positively to high-quality core instruction and tier 2 interventions (e.g., Burns et al. 2005; Torgesen 2002; Vaughn et al. 2003). If, however, empirically supported tier 2 interventions are implemented with fidelity yet the individual does not favorably respond to these interventions, then an increased intensity and duration of intervention is provided in tier 3 levels of support. RTI logic suggests if a student is exposed to multiple tiers of interventions and still does not develop essential academic, social, or behavioral skills, the student may have a disability that warrants entitlement to special education services (Barnett et al. 2004; Duhon et al. 2004; Kovaleski et al. 2013; VanDerHeyden et al. 2003).

The RTI framework is built around titrating instruction and intervention based on a student's academic, social, or behavioral needs. Consequently, a key feature of the process is the use of objective data to evaluate effectiveness of instruction and intervention offered across the tiers of

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_8

support. Put another way, RTI demands that data are regularly collected, reviewed, and interpreted to determine if the learner is responding to intervention (National Association of State Directors of Special Education 2005; Reschly et al. 1999). This feature of RTI can be applied to a particular student, a small group of students (e.g., class or grade level), or an entire student population (e.g., school or district). Data from multiple sources are typically analyzed when evaluating progress including grades, anecdotal reports, universal screenings, and data from progress-monitoring tools. These latter two groups of data have received considerable empirical and practical attention, given their impressive psychometric qualities and efficiency (Brown-Chidsey and Steege 2005; Fuchs 2003), although Christ and his colleagues have recently reported data on psychometric issues associated with curriculum-based measurement (CBM) that should be taken into account by practitioners (Christ et al. 2012; Christ et al. in press).

Fuchs and Deno (1991) described general outcome measures (GOMs) as instruments that reliably assess students' skill acquisition within a particular curricular domain. The greatest amount of attention regarding GOMs, and, in particular, universal screenings and progress-monitoring tools employed within an RTI framework, has been in the area of reading. The earliest measures of reading development focused on oral reading fluency (ORF) and preliteracy skills such as phonemic awareness, rapid serial naming, and letter naming. In fact, a whole industry of universal screenings and progress-monitoring tools has emerged in the past 15 years that has capitalized on the overwhelming body of evidence indicating that these skills are vital to reading comprehension and academic success (e.g., National Reading Panel 2000; Shinn 1989). Measures of reading comprehension, in particular, cloze reading tasks, have emerged in recent years to fill the void of a reading comprehension fluency measure. A cloze-reading procedure requires the individual to silently read a passage in which every *n*th (e.g., seventh) is replaced with three words from which the reader must select the correct option (see Gellert and Elbro 2013 for a review). Two of the more popular cloze procedure tasks are Dynamic Indicators of Basic Early Literacy Skills (DIBELS Next Daze; Good et al. 2011) and AIMSweb Maze (NCS Pearson 2011). In addition, a number of computer-adapted screening tools based on item response theory that assess reading comprehension have also been developed, such as STAR Reading (Renaissance Learning 1996) and the measures of academic progress (MAP; Northwest Evaluation Association 2004). Schools that implement RTI typically establish an annual-assessment plan that includes systematic universal screening and progress monitoring of preliteracy skills and ORF in the primary grades and ORF and reading comprehension in the intermediate grades.

Similar trends are observed in the areas of writing and mathematics, although the corpus of empirical support for these tools is far less robust than in reading (Kelley 2008). Measures of writing skill acquisition include those that monitor the fluency of correct letter sequences in spelled words at the primary grades and, at the intermediate grades, fluency of correct word sequences. Within the mathematics domain, fluency in the early numeracy skills of seriation, numeration, and number identification are typically assessed in the primary grades. Assessment of mathematical computation fluency and problem-solving occurs in the intermediate grades.

RTI is traditionally known for its emphasis on academic skill development. The other side of the RTI pyramid is RTI for behavior, known as positive behavioral interventions and supports (PBIS; Sugai and Horner 2009). Data that are typically monitored at tier 1 in school-wide PBIS (SWPBIS) include office discipline referrals (ODRs), positive student acknowledgements, student attendance and tardies, and surveys of organizational health and beliefs about school safety and culture (Algozzine et al. 2010). At tier 2 levels of support, interventions such as check-in/check-out (CICO) are implemented with daily points earned for positive behavior in addition to ODRs, attendance, and academic performance (Crone et al. 2010).

Accounting for some regional or local differences in the GOM data collected within an RTI or PBIS framework, it is reasonable to conclude that most schools endorsing these multi-tiered service delivery models usually collect many of the data previously reviewed. Collecting such data, however, is only the beginning of the data-analysis process. Next, schools must systematically review and interpret these data in the context of the instruction and interventions implemented with students.

Data-Analysis Teaming, RTI, and PBIS Systems

By choosing to implement RTI and/or PBIS, schools necessarily embrace the premise that data should guide their efforts to improve students' academic achievement, facilitate students' emotional and behavioral well-being, and evaluate the efficacy of their programs. The collection of the many and varied sources of data is worthless without the careful and structured analysis of these data. In particular, data need to be analyzed at the individual, grade, building, and district level to make decisions about individuals and groups of students as well as district-wide policy decisions. At the building level, Kovaleski and Pedersen (2008; 2014) have proposed that DATs should be empaneled to meet regularly to review student data and make instructional and management decisions. They have also developed forms for use by DATs to guide the meetings and annotate student data and team decisions.

DAT Membership Analysis of student data should primarily occur at grade-level meetings of teachers and specialists who have students in common. Although some secondary schools have reported success with department-level meetings, grade-level clusters are preferable in most situations (Kovaleski and Pedersen 2008). Other key DAT members are the administrator in charge (typically the school principal) and a person who manages the data.

Meeting Format DATs use the problem-solving process (Tilly 2008) to review data, set goals, select strategies, plan for logistics of implementation, and monitor fidelity of strategy implementation. At tier 1 of the multi-tier (RTI) process, the DAT examines data on all students in the grade level. For academics, the DAT analyzes universal screening data to ascertain the percentages of students scoring in proficient, emerging, and at-risk ranges (usually using terminology associated with the particular assessment tool). For behavior, the DAT reviews class-wide or school-wide data on types and locations of discipline infractions. From these data, goals are set for the group as a whole, and strategies that will be implemented by all teachers devised. An important aspect of tier 1 data analysis is that the DAT does not review data on individual students at this point because that analysis leads the group away from the consideration of class-wide and school-wide implementation. The DAT also plans for how the instruction or management strategies will be supported (e.g., through peer coaching) and how treatment integrity will be accomplished (e.g., through the observation by the principal).

At tier 2, the DAT reviews data on individual students and creates groups of students with similar needs for supplemental intervention planning. For example, the DAT might use the results of universal screening in reading to identify fourth graders for intervention groups that differentially target comprehension versus fluency versus phonics skills. In general, interventions for these groups follow a *standard-protocol* format that features structured, manualized approaches that are based on scientific research (Vaughn and Fuchs 2003). For behavior, students who display particular types of behavioral difficulties might also be provided with CICO or a group intervention (e.g., social skills building). The DAT would also plan for ongoing progress monitoring of all students receiving tier 2 support.

For tier 3, the DAT would again review data on individual students, particularly data from progress monitoring and any further assessments that might be conducted. When students fail to make progress on the basis of tier 2 interventions, many DATs use more extensive assessment procedures to identify students' individual needs. For example, curriculum-based evaluation (CBE) procedures might be used to identify specific deficiencies in academic skills (Howell and Nolet 2000), while functional behavioral assessment (FBA) would be used to develop hypotheses about possible antecedents and maintaining consequences of the problem behavior (Steege and Watson 2009). These data would be used to craft customized interventions for the individual students.

Data Warehousing

An important consideration for any district choosing to implement an accountability system within an RTI framework is deciding how to store the data, a natural outcome of an effective RTI initiative. There are myriad options on the market for this purpose, ranging from very reasonable paper documents to extremely complex electronic solutions. Regardless of the final decision, all districts must utilize a system that not only works within their personnel, financial, and technological constraints but also is easily accessible, reliable, and facilitates measurement of important outcomes. While districts typically begin with a paper solution, many will ultimately transition to a more complex electronic solution, possibly beginning with a spreadsheet but likely resulting in a contract with a software company that allows for server-based storage of data. Some districts refer to this electronic solution as a *data warehouse*. This is essentially a storage option where educational data from several areas that affect student learning are housed. These data then can be used to produce a variety of reports for analysis and review by a wide range of audiences, from students to staff to parents.

A clear advantage of an electronic server-based data warehouse solution to data storage is that all data become available and accessible to multiple staff members and across a network. Even better is a web-based server solution to data storage that allows for access to data outside of a school district's network. This option allows a staff member to access records while away from the office or classroom, while allowing access to records by parents or even students, should the district provide for such access. Regardless of the solution, the district must consider security and confidentiality of the data. Security can be achieved in electronic platforms through password-protected and/or restricted-access features; paper-based solutions may be less secure and only accessible while on campus.

Data Warehousing Considerations An effective electronic data warehouse will allow for districts to do more than just store data. It will permit districts to explore and investigate or analyze data, especially through preset queries, filters, or report formats that allow for score reports, visual displays of performance, access to details of student performance across time and settings, etc. A more powerful system will allow for self-defined queries so that districts may choose to investigate how a particular cohort compares to previous cohorts or how students receiving a particular intervention compare to students without intervention or in a different intervention on the same benchmark, as measured by progress-monitoring data, a GOM, or a large-scale assessment. Districts should also consider whether the queries may be saved for future analysis, which may be a considerable time saver if districts find themselves generating the same reports each time they review data, whether it is with different cohorts of students or even if tracking the same cohort across time.

A strong electronic solution will also allow districts to extract data and present it in a visual display that is easily understood. This may include the use of bar charts, pie charts, scatter plots, histograms, and time-series graphs. This feature of electronic data warehousing allows for easy and efficient data exploration or investigation from large groups to individuals within one screen. A consideration when selecting a data warehouse is the flexibility of the program and whether it permits users to alter labels, modify time periods for extracting certain data, or identify intervention periods. Availability of these features will assist teams when analyzing data and making data-informed decisions.

An electronic data warehouse allows districts to track and monitor students and groups over time, short- and long-term, and across multiple delivery or intervention options. In addition, a strong solution will also allow districts to conduct trend analyses to efficiently make informed, fact-based decisions about instruction and intervention. This analysis may be applied to individuals as well as groups of students and teachers.

Most advantageous is a system that allows for districts to calculate trend lines and aim lines that may assist schools in determining the effectiveness of an intervention based on the trajectory of growth, particularly when compared to a normative sample. Districts may also want to consider selecting a system that allows for the documentation of intervention efforts or changes in interventions, which will help determine if a particular intervention produces more desirable effects or outcomes. Considerations should include the ability to compare student to self, student to other students, groups of students to groups of students, grade to grade, teacher to teacher, school to school, etc.

Since no system will serve all purposes, it is important for districts to consider the ability of data warehousing systems to interact with or communicate with other data warehousing systems used by a district. For example, a district may utilize one electronic solution for demographics, another for housing its academic and behavioral data, and yet another for creating special education documents such as evaluations, eligibility reports, or individualized education programs (IEPs). Districts must consider a data warehouse that allows for two-way seamless communication with other systems so that data from one system may be uploaded or downloaded into another. This integrated approach not only decreases the likelihood of duplicated efforts but also ensures the accuracy of data that will be analyzed. In addition, one system may allow for a different type of data analysis or presentation not offered by another system. Integration with special education paperwork systems also ensures that the assessment data gathered as part of an RTI framework is utilized as part of eligibility decision-making and special education documentation.

As districts decide which electronic data warehouse solution will best meet their needs, there are some strengths and limitations of data warehousing to consider. The positives of an electronic data warehouse may outweigh the negatives, particularly if a district performs a thorough analysis of its needs and carefully chooses a solution. It is critical that when selecting a solution, districts solicit input from stakeholders from every level in the education system. Stakeholder involvement in the exploration and investigation of the warehouse is needed to educate all who might be affected by a data warehouse regarding what a data warehouse can do for the district and how it can advance learning. Stakeholder education will also facilitate buy-in, acceptance of the data warehouse as a tool in the decision-making process, and create a vision of the use of data to accelerate learning for all students. Asking stakeholders (e.g., administrators, teachers, school counselors, school psychologists) about their specific data needs and desires helps to inform the purchasing decision.

Identifying Nonresponders Ultimately, the data warehouse(s) should facilitate a secure, efficient, and reliable process by which data are entered, stored, and extracted in a meaningful and helpful manner. The DATs, in turn, would have multiple data available to make appropriate, evidenced-based decisions about instruction and intervention. These decisions should be helpful to DATs focused on the progress of individual students, typically via a single-subject design methodology and visual inspection of a time-series graph. Such venues allow teams to determine if intervention efforts are having the desired effect for a particular student. Data warehousing can also assist in the decision-making process regarding instruction and intervention effects for small groups such as classrooms or entire grade levels.

Students who continue to display deficiencies and fail to make meaningful progress in spite of significant intervention attempts across three tiers of support should be considered for special education eligibility. Kovaleski et al. (2013) have outlined extensive procedures for harvesting data developed during the provision of tiered support to inform the eligibility decision-making process. The data management structures described in this chapter have special applications for the efficient collection of these data.

Evaluating Programs Practitioners using RTI and PBIS expend significant efforts to facilitate and enhance effective instructional and management practices in tier 1 and to design, implement, and monitor interventions for students receiving supports in tiers 2 and 3. Periodically, it is critical for schools to “take a step back” and analyze whether these efforts are producing the expected outcomes for students as a group and for individual students who display difficulties. Building-based and district DATs are recommended for accessing and analyzing data sets to conduct formal evaluations of these programs. Again, the utility of data warehousing solutions to facilitate these evaluations is compelling.

Considerably more sophisticated statistical procedures may be applied, given typically larger numbers of students included in these analyses. Many electronic solutions have built-in statistical packages for conducting various descriptive and inferential statistics. At a district, regional, or statewide level, these data warehouses can influence the quality, quantity, and type of data reviewed in making decisions that have a great effect on large numbers of students, parents, educators, and communities. For example, PBIS scale-up efforts are greatly enhanced by cross-district data warehousing and reports generated from aggregated data across multiple school districts (Illinois PBIS Network 2012; Runge et al. 2012). Reports such as these can provide the empirical support needed to drive policy and practice on a large scale. Access to efficient and effective data warehousing systems will ultimately allow these decisions to be made with reliable and valid data.

Data Warehousing Examples

A number of commercial data warehousing systems are available to assist in the entry, storage, and retrieval of data used within RTI and PBIS frameworks. Some of these systems are components of a particular set of assessment products (e.g., DIBELS, AIMSweb, STAR, MAP), while others are systems that are independent of assessment products but aggregate data from multiple sources into one central repository (e.g., Performance Tracker). Although a comprehensive review of all data warehousing systems is not possible within this chapter, a few of the more established data warehousing systems are presented below. First, a brief review of system features is provided, then a sample of reports or visual displays is offered. Finally, a few notable strengths and limitations are indicated. The purpose of these reviews is not to provide a comprehensive, fully vetted critique of the systems, but rather to provide an introduction to the features available. These reviews are not to be construed as product endorsements.

RTI A number of web-based data warehouse products are available that store data specific to a particular assessment line. Two of the more popular systems include DIBELS Next (http://dibels.org/next.html) and AIMSweb (http://www.aimsweb.com/). A review of these systems, however, is not provided here, given most readers' familiarity with these products. Other comparable products include STAR Assessments (http://www.aimsweb.com/) and MAP (http://www.nwea.org/). The list of available systems that track benchmarking and progress-monitoring data for RTI seemingly expands each year. For the most part, these systems have many similarities regarding data entry, storage, report generation, and visual displays of data albeit using a different assessment product connected to its data warehouse.

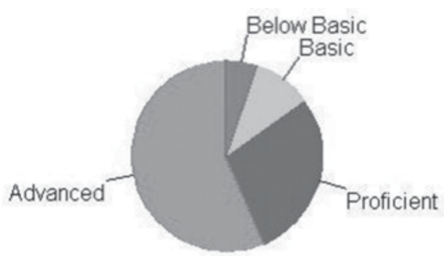
While many are likely familiar with one or more of the data warehousing systems indicated above, newer data warehousing systems independent of specific assessment products are gaining popularity in the RTI process. These systems pull data from various websites, including AIMSweb or DIBELS, for example, and combine these data with other data such as attendance, grades, and discipline data. As a consequence, these systems tend to function as repositories of data gleaned from other warehouses so that educators can access multiple data in one central location. A growing number of these products are appearing, with two being familiar to these authors: Performance Tracker (https://butler-pa.perfplusk12.com/Default.asp) and Comprehensive Data analysis (CDA; https://cda.aiu3.net/cdanew/default.shtm). Due to the similarity of these products, only Performance Tracker is reviewed here.

Performance Tracker

Performance Tracker is a component of PerformancePLUS, SunGard's web-based, districtwide system that assists educators in tracking and analyzing student performance against state standards, mapping and managing curriculum, as well as building and administering local on-line benchmark assessments. Within Performance Tracker, educators access one location to upload, enter, store, review, and monitor student data in longitudinal format, thus decision-making becomes extremely efficient. These decisions should then enhance instruction and increase both individual student achievement as well as overall district achievement. Assessment data may be aggregated or disaggregated in a variety of preset or user-determined ways through myriad filters or queries. User-specific report queries may be saved for future reference or use. All data may be presented in a variety of interactive presentation formats including charts and graphs. Moreover, all data may be exported into Microsoft Excel™ files to facilitate detailed analyses and presentations not available within the on-line warehouse.

Among other things, reports may be generated to analyze performance on large-scale assessments to view students by performance level (see Fig. 1), to compare performance across time on assessments from benchmark to GOM (see Fig. 2), to analyze performance on individual items on assessments (see Fig. 3), or to download results from single or multiple assessments into a user-friendly comma-separated value (CSV) format or Microsoft Excel™ spreadsheet (see Fig. 4). Performance Tracker will accept data for manual entry into preset fields or ones that are fully customizable. For example, preset fields are built into the system for DIBELS, AIMSweb, Advanced Placement, SAT, American College Testing (ACT), large-scale state assessments of the state standards, etc. Districts may create their own fields for local benchmark assessments and diagnostic assessments. In addition, data may be uploaded into Performance Tracker through a Microsoft Excel™ spreadsheet, and all data will seamlessly upload into preset fields. While Performance Tracker is helpful for storing data and running reports to use for data analysis, the system does not provide for any statistical analysis of the data.

PBIS A number of secure, on-line data warehouses exist for schools implementing PBIS. The School-Wide Information System (SWIS; http://swis.org; Educational and Community Supports 2011) is a database system that tracks behavioral and intervention data at all three levels of PBIS implementation. The PBISAssessment.org website provides the capacity to collect and analyze staff survey and PBIS implementation integrity data for individual schools. The PBISEval.org website provides licensed administrators the ability to aggregate data from SWIS and PBISAssessment and monitor and evaluate data trends across multiple schools and districts. Unrelated to the aforementioned family of on-line secure websites is AIMSweb Behavior (http://aimsweb.com/behavior), a comprehensive system of tracking universal screening and progress-monitoring data and repository of intervention modules linked directly to the data. Lastly, Tracking Referrals



85.10% Proficient

Proficiency Level	# of Tests	# of Students	Percent	View Students
Advanced	1	144	56.47%	Click to view the students
Proficient	1	73	28.63%	Click to view the students
Basic	1	25	9.80%	Click to view the students
Below Basic	1	13	5.10%	Click to view the students

Fig. 1 Sample pie chart showing students by performance level

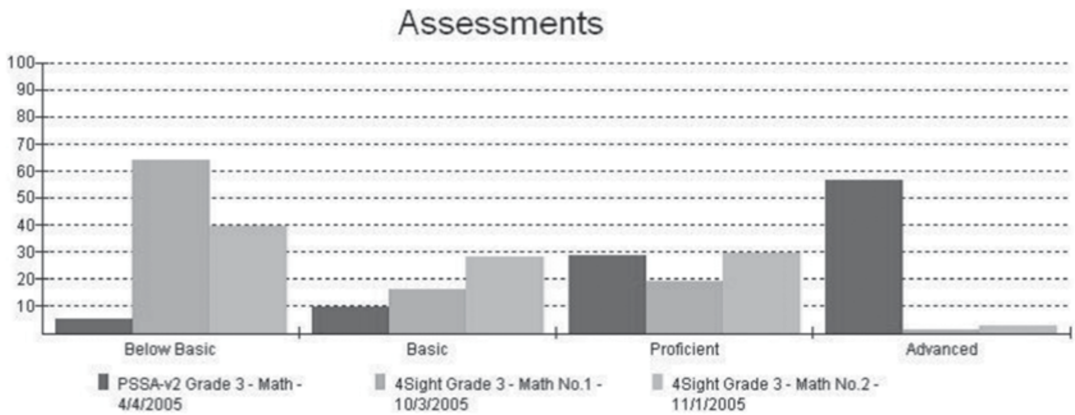


Fig. 2 Sample histogram showing comparison on benchmark assessments

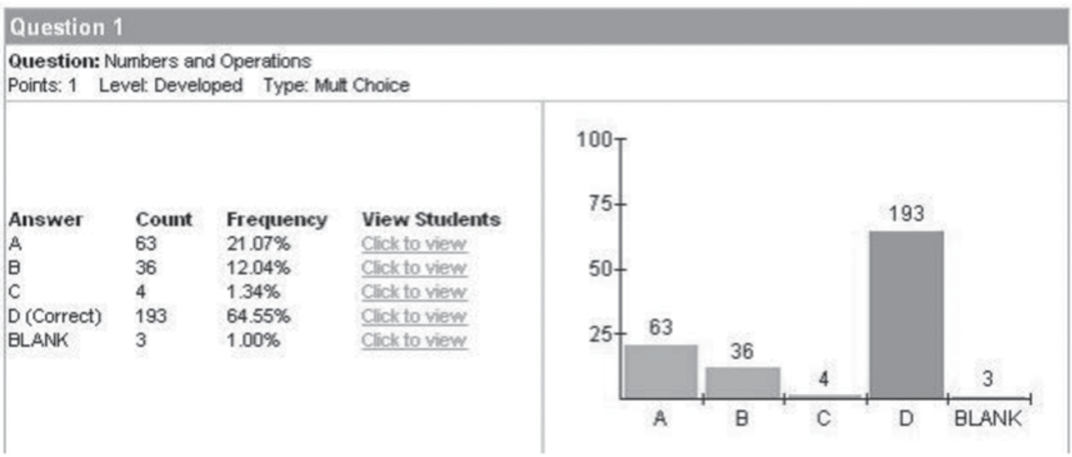


Fig. 3 Sample graph of assessment item analysis

Encouragements Notifications Discipline Safety (TRENDS; http://www.pacificnwpublish.com/trends/index.html) is a relatively new system that is comparable to SWIS with the addition of a capacity to track positive commendations and school safety data. Highlights of these systems are offered below.

SWIS

Within the SWIS system, data about individual ODRs are efficiently entered by a designated staff person with administrative privileges to the site. Data entered include the name of the student, sex, grade, special education status, time of infraction, location of the event, concerning behavior, perceived motivation for the behavior, information about others who may have been involved, referring staff person, and resultant administrative action taken. These data are maintained on the secure website to which those with *read-only* access can log in and view individual ODR reports or generate a variety of reports or visual displays.

The primary strength of SWIS is the ease with which data are entered and the range of data reports generated. The SWIS system has five standard ODR graphs that school teams are encouraged to review monthly. These *Big 5* are average number of ODRs per day per month, ODRs by problem behavior, ODRs by location, ODRs by

Subject: Math
Grade: Eighth Grade
Filter: Assessments: Math - Grade 8 - Qtr 1 10/25/2005

Student Name	Student ID	Math - Grade 8 - Qtr 1-M8.A.2.1	PL	Math - Grade 8 - Qtr 1-M8.C.3.1 10/25/2005	PL	Math - Grade 8 - Qtr 1-M8.D.1.1	PL	Math - Grade 8 - Qtr 1-M8.D.2.1	PL	Math - Grade 8 - Qtr 1-M8.D.2.2	PL	Math - Grade 8 -	PL
	110049	4	Adv	4	Adv	4	Adv	4	Adv	3	Pro	19	Adv
	110129	4	Adv	4	Adv	3	Pro	4	Adv	4	Adv	19	Adv
	110128	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	110048	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	171743299	4	Adv	3	Pro	4	Adv	4	Adv	3	Pro	18	Adv
	110120	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	110163	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	178741816	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	190740490	4	Adv	4	Adv	2	Bas	4	Adv	2	Bas	16	Pro
	186746492	4	Adv	3	Pro	4	Adv	3	Pro	4	Adv	18	Adv
	110116	4	Adv	4	Adv	3	Pro	4	Adv	4	Adv	19	Adv
	110041	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	110169	4	Adv	4	Adv	3	Pro	3	Pro	4	Adv	18	Adv
	110171	4	Adv	4	Adv	4	Adv	3	Pro	4	Adv	19	Adv
	110039	4	Adv	3	Pro	4	Adv	4	Adv	3	Pro	18	Adv
	110196	4	Adv	4	Adv	4	Adv	2	Bas	4	Adv	18	Adv
	110186	2	Bas	3	Pro	4	Adv	4	Adv	3	Pro	16	Pro
	110034	4	Adv	3	Pro	4	Adv	2	Bas	4	Adv	17	Pro
	110185	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv
	110173	4	Adv	4	Adv	4	Adv	4	Adv	4	Adv	20	Adv

Fig. 4 Sample of downloadable spreadsheet

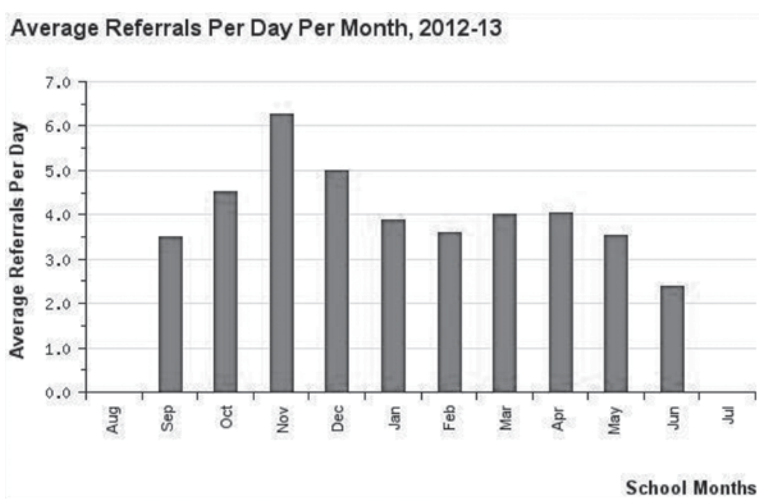
time, and ODRs by student. A sample graph is presented in Fig. 5 representing an efficient view of daily ODR trends by month. These data help DATs determine the relative efficacy of the SW-PBIS framework from month to month and make modifications to interventions based on these screening data.

Other preestablished reports and graphs available in the SWIS system include suspension/expulsion data, ODRs by ethnicity, year-end reports, ODRs by referring staff member, and annual *triangle data* report. This latter report provides DATs with the percentage of the student population receiving 0–1, 2–5, and 6+ODRs in an academic year (see Fig. 6 for an example). Spaulding et al. (2010) validated these three ODR ranges to represent approximately 80%, 15%, and 5%, respectively, of a typical school population. Therefore, the triangle data graphs help school teams evaluate the efficacy of the SWPBIS framework across the entire student population relative to national trends in SWPBIS efficacy. A nearly limitless array of custom graphs and reports can also be generated within SWIS by selecting out certain criteria.

Three limitations to the SWIS ODR system need to be mentioned so that schools using this system can make adjustments to their data analysis plans. First, with the exception of the first of the Big 5 (i.e., average number of ODRs per day per month), the remaining graphing options do not allow for an easy inspection of trends across different months. For example, suppose a school DAT recognized a spike in ODRs in the cafeteria during the month of October. One relatively simple intervention would be to reteach cafeteria behavior in the first week of November. At the early December DAT meeting, a review of ODRs in the cafeteria should include analysis of trends in October compared to November when the booster/reteaching sessions were completed. Such a longitudinal comparison of ODR trends is not easily accomplished within the current version of SWIS.

A second limitation with SWIS relates to the Ethnicity Report which provides the percentages of ODRs by ethnicity. Given results from Skiba and colleagues (Skiba et al. 2002; Skiba et al. 2011) indicating Blacks and Hispanics are disproportionately disciplined compared to their White counterparts, it is imperative that schools are sensitized to their own practices by continually monitoring disciplinary practices toward minority groups. At present, SWIS calculates risk indices for each ethnicity group. The risk index in the context of ODRs is the proportion of a group that is at risk for receiving an ODR. These data have limited interpretability, however, in the absence of a comparison group. Hosp and Reschly (2003) recommend calculating a risk ratio which compares the relative-risk index of one group to the risk index of another group. Risk ratios

Fig. 5 Sample graph of SWIS average referrals per day per month



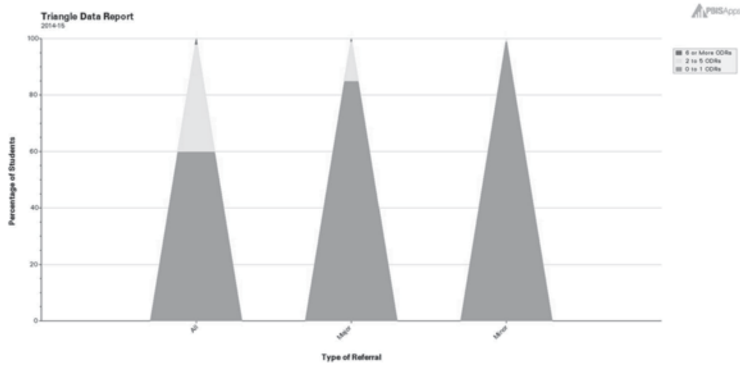


Fig. 6 Sample SWIS triangle data report

provide more meaningful data about whether particular groups are being disproportionately disciplined relative to other groups. Unfortunately, the SWIS system does not provide risk ratios, so these critical data need to be calculated by hand. Moreover, school teams are also encouraged to review potential disproportionality among other subgroups based on socioeconomic status, special education eligibility, and sex despite the fact that SWIS does not generate these types of reports (Boneshefski and Runge, 2014). A third limitation to SWIS is that it only tracks ODR data at the universal level. SWIS does not track data regarding the acknowledgment system that is an essential feature of SWPBIS nor does it track any other universal screening data.

At the secondary tier of PBIS support, a standard protocol intervention is CICO. The SWIS system enables schools to enter data about particular students enrolled in this intervention. As interventions are implemented with a student, the SWIS–CICO system allows teams to indicate such changes to the behavior plan, which then appear on the generated time-series graphs. This feature allows for efficient visual analysis of intervention effects using widely adopted applied behavior analytic techniques (Cooper et al. 2007). A sample graph is displayed in Fig. 7 and includes two phase changes, represented by solid vertical lines, on September 10 and 19. As with the ODR reports, there are a handful of different graphing options available for SWIS–CICO.

A final component to SWIS that recently became available is the individual student intervention system (ISIS), which serves as a repository of data and important documents for students at a tier 3 level of support. For example, ISIS will store functional behavioral assessment and positive behavior support plan (PBSP) documents, data from PBSP implementation efforts, and mechanisms to track which team members are responsible for which tasks related to the development, implementation, and monitoring of the PBSP.

PBISAssessment.org

The PBISAssessment.org website is a central repository for collecting and analyzing staff survey and fidelity of implementation data. Among the staff surveys available on this website are the Effective Behavior Support: Self-Assessment Survey (EBS: SAS; Sugai et al. 2009) and School Safety Survey (SSS; Sprague et al. 2002). The former survey assesses staff perceptions about the quality and need for improvement in behavioral support. The SSS is a survey of staff perceptions of risk factors associated with school violence and protective qualities within and outside the school community. Four fidelity measures are also available on the PBISAssessment.org website. The School-wide Evaluation Tool (SET; Sugai et al. 2005), Benchmarks of Quality (BoQ; Kincaid et al. 2005), and Effective Behavior Support: Team Implementation Checklist (TIC; Sugai et al. 2009) are the traditional instruments used to measure fidelity of SWPBIS. Fidelity of tiers 2 and 3 implementation is documented using the Benchmarks for Advanced Tiers (BAT; Anderson et al. 2012). Lastly, the recently developed fidelity measure for preschool and early childhood program-wide PBIS, the Early Childhood Benchmarks of Quality (EC BoQ; Fox et al. 2010), is available on this website.

PBISAssessment.org allows DATs to coordinate electronic staff surveys and warehouses data in a manner that facilitates quick report generation for review. A sample 3-year review of SSS data is offered in Fig. 8. Additionally, annual and longitudinal reviews of fidelity data are quickly accomplished within this system.

PBISEval.org

PBISEval.org is the third member in the SWIS and PBISAssessment.org family of data warehouse systems available for schools implementing PBIS. The PBISEval.org website allows district- and state-level PBIS coordinators to monitor fidelity status, ODR and CICO data, and staff survey results from multiple schools at one time. This website has particular utility for those interested in scaling-up PBIS efforts, while readily accessing data to help problem solve and make informed decisions.

A series of guiding questions are offered on this website to help users aggregate and disaggregate data from the SWIS and PBISAssessment.org websites. Figure 9 provides an example of aggregated BoQ fidelity data across a 3-year pe-

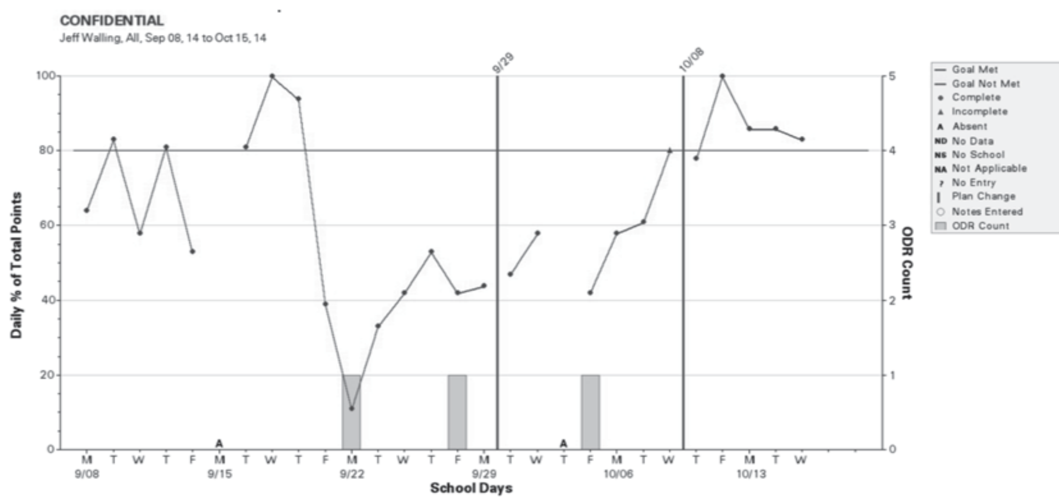


Fig. 7 Sample SWIS–CICO individual student count report

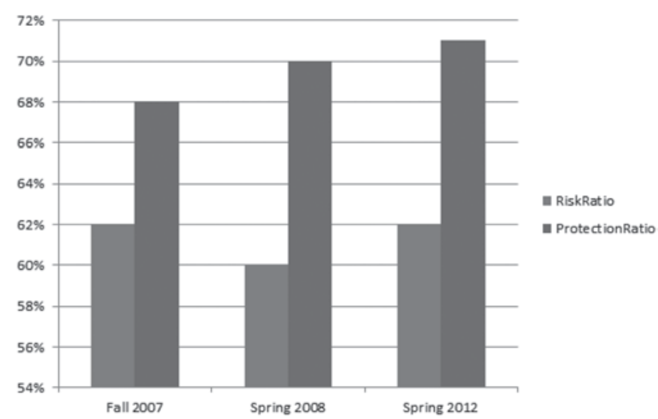


Fig. 8 Sample 3-year review of school safety survey results

riod disaggregated by building level. This visual display facilitates an efficient view of the number of schools, by building level, which achieved the minimum threshold for designation as fully implementing SWPBIS.

AIMSweb Behavior

AIMSweb Behavior is a comprehensive system that includes data warehousing and intervention modules for children from preschool to 12th grade. A review of the intervention modules is beyond the scope of this chapter; however, readers should take note of this feature of AIMSweb Behavior which far exceeds what many other PBIS data warehousing systems provide.

The universal screening data entered into AIMSweb Behavior include teachers', parents' and, when appropriate, students' ratings on the Behavioral Assessment System for Children-2, Behavioral and Emotional Screening System (BESS; Kamphaus and Reynolds 2007). The BESS data are entered into the secure website and various reports can be generated for DATs to review. Additionally, the two broad domains from the Social Skills Improvement System Performance Screening Guide (SSIS-PSG; Elliott and Gresham 2007) are completed by teachers for their entire class roster: prosocial behaviors and motivation to learn. These ratings are completed using a behaviorally anchored Likert rating from 1 to 5.

Once data are entered into the AIMSweb system, various reports can be generated for individual students, classrooms, grades, and entire buildings. Primarily, these reports highlight student levels of risk for significant behavioral or social dysfunction based on data from the BESS and SSIS-PSG. Once students are identified for tiered levels of support via the universal screening data, the AIMSweb Behavior system provides a wide menu of empirically validated interventions to address skill and/or performance

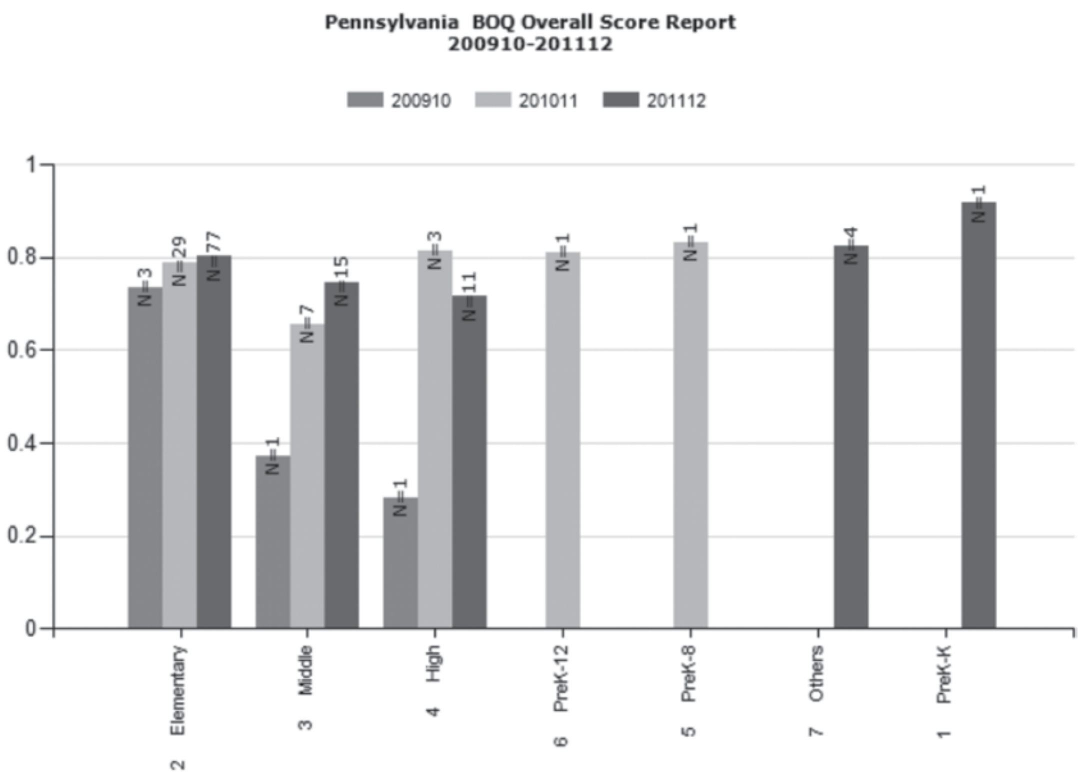


Fig. 9 A 3-year BoQ fidelity data aggregated across Pennsylvania

deficits. AIMSweb Behavior then facilitates the efficient tracking of progress-monitoring data and production of single-subject design graphs which are used to evaluate intervention efficacy. See Fig. 10 for an example of a Progress-Monitoring Improvement Report regarding overall behavioral points earned by day. Additional reports allow for drilling down to daily points earned by specific behaviors targeted for intervention.

Advantages of AIMSweb Behavior include tracking universal screening data other than ODRs. Collecting and analyzing teacher, student, and parent ratings of social and emotional functioning serves as a much more proactive and preventative method of identifying risk and implementing early intervention efforts before student's behavior intensifies to a magnitude warranting an ODR. The built-in intervention modules are immensely helpful for DATs as they design, implement, and monitor intervention efficacy. Still another advantage of AIMSweb is that it is one of the few systems that allows for warehousing of data pertinent to both RTI and PBIS. A limitation of AIMSweb Behavior, however, is that fidelity of PBIS implementation is not accomplished via this system. So although a single school could monitor PBIS fidelity using paper–pencil versions of BoQ, SET, or TIC, scale-up efforts in which multiple schools are implementing PBIS would necessitate utilization of the PBISAssessment.org website in addition to AIMSweb Behavior.

TRENDS The TRENDS warehouse system provides many similar features as the SWIS system in regard to tracking ODRs by day of week, race/ethnicity, sex, time of day, location, and problem behavior. Like SWIS, TRENDS has standard reports and graphs that are created from the ODR data; however, TRENDS has more flexibility in generating customized graphs. Additionally, TRENDS allows for an analysis of whether disproportionate disciplinary practices are occurring in a building in a more sophisticated and technically precise manner than SWIS by providing additional options for data analysis built into the system itself.

A feature of TRENDS not found in any other data warehousing system is the capacity to enter and monitor data regarding commendations—the reinforcement systems that schools implementing SWPBIS typically utilize. Tracking the acknowledgement of prosocial behavior is important for DATs to monitor, given that the level of problematic behavior and prosocial behavior is likely wedded to the schedule of reinforcement. TRENDS also warehouses data from electronically completed teacher, parent, and student surveys of climate and safety so that these data can be incorporated into the traditional ODR and

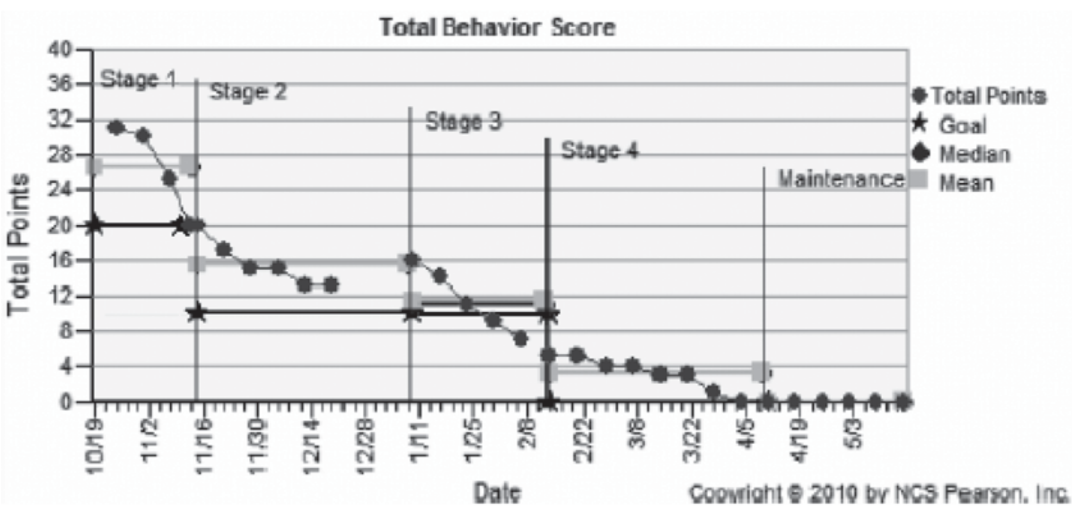


Fig. 10 Sample AIMSweb progress-monitoring improvement report

commendations data to provide a broader picture of SWPBIS efficacy.

At tiers 2 and 3, TRENDS, similar to AIMSweb Behavior, has a library of empirically derived behavioral interventions from which PBSPs can be developed. Once the PBSP has been developed, TRENDS facilitates efficient storage of daily behavioral data and generates a variety of graphs used in the evaluation of intervention effectiveness. All of these features, similar to the BESS and SWIS, allow DATs to be more efficient in their data analysis and documentation of intervention efforts and efficacy.

Areas for Future Research

Integrated systems that warehouse RTI and PBIS data will be particularly useful as more districts simultaneously implement both initiatives. AIMSweb provides such an integrated RTI–PBIS data warehouse system with its own built-in assessments; however, more systems such as this are needed. Therefore, it will be important to develop warehousing systems with increased flexibility to store other assessment products and/or aggregate data from multiple websites. Expansion of aggregated data warehousing systems will make accessing multiple data for RTI and PBIS considerably more efficient.

DAT meetings have been occasionally observed in which team members become overwhelmed with the volume of data available. With the expanding quantity and quality of data available through data warehousing systems, future research must focus on how to make DAT meetings more efficient. An empirically based teaming protocol, such as that offered by Todd et al. (2011) for SWPBIS, is a starting point for researchers to validate a similar structure to help RTI DATs be productive and efficient in maximizing available data. The forms and process proposed by Kovaleski and Pedersen (2008; 2014) would be important to include in that line of research.

Cross-system data warehousing capabilities (e.g., www.PBSEval.org) will facilitate future research on large-scale program evaluations of critical outcomes associated with RTI and PBIS. This will not only aid in a district-level review of RTI or PBIS implementation but also provide the vehicle for cross-district, statewide, or national efficacy evaluations. These evaluations will provide the empirical evidence and practical implications to sustain and expand both initiatives. To some extent, these latter evaluations are emerging in the PBIS literature (e.g., Eber et al. 2010; Bradshaw et al. 2010; Runge et al. 2012); however, considerably more work must be completed in both RTI and PBIS.

Warehousing data across buildings and districts within a particular region allow for the development of local norms. Stewart and Silberglitt (2008) suggested that the availability of local norms will augment the decision-making process as a means of reducing bias (Marston et al. 2003) and provide instructional feedback at the micro- and macro-levels (Howe et al. 2003). As systems become omnipresent, individual and regional school districts can aggregate their data to formulate these local or regional norms. Moreover, the increased number of students' data contained in cross-system warehouses may provide a large enough number of students in various subgroups (e.g., IEP, economically disadvantaged; racial groups) for local and regional norms to be developed for these populations.

Implications for Practice

As schools and districts move forward with RTI and/or PBIS implementation efforts, careful consideration of the data warehousing systems critical to making data-based decisions is necessary. A summary of broad considerations within four broad domains is offered in Table 1: selection committee, selection process, system features, and system use. The suggested considerations are not intended to be exhaustive but should provide initial discussion points to guide the selection of a data warehouse system that adequately meets the needs of all stakeholder groups and maximizes the potential of the data warehousing system eventually chosen.

Table 1 Data warehousing systems—implications for practice

Domain	Broad considerations
Selection committee	Are all stakeholder groups represented?
	What are stakeholder groups' interests and concern?
Selection process	What is the range of purposes for the data warehousing system?
	What data need to be collect?
	Does the system meet the needs of all stakeholders?
	Is a web- or server-based system required?
	Cost?
	What are the training and technical support needs?
System features	How are the security and confidentiality of data maintained?
	What kind of data are entered?
	How much flexibility is there in adding/changing data entry fields?
	Is there a mechanism to enter fidelity data?
	What level of access to the system do we want?
	What is the range of preset queries we would like?
	Are customized queries available?
	Are customized queries saved for future use?
	What type of reports and visual displays are available?
	Alterability of reports, custom reports, and visual displays?
	Can aim line, trend lines, and normative comparison data be included in reports and visual displays?
	Can data be downloaded?
	In what format?
	Compatibility with other data warehouse systems?
System use	What is the data entry process/procedure?
	Who is responsible?
	Time line for data entry?
	Resources to efficiently enter data?
	Who has full versus read-only access?
	How is the accuracy of data ensured?
	What is the data retrieval process/procedure?
	Who is responsible?
	What are the guiding questions used to retrieve relevant data?
	How and when will data be disseminated to DAT?
	How are the data reviewed?
	Who has access to review?
	When should they review?
	What are the guiding questions to productively team around data?

DAT data analysis team

First, a steering committee should be formed that includes members from key stakeholder groups so that each group's interests and concerns are considered. Next, the committee should come to consensus on the purposes of the data warehousing system so that decisions regarding the kinds of data to be collected, accessibility of data, and training and technical support needs can be made. Within those discussions should be consideration of the various features of the data warehousing system, including an analysis of strengths and limitations of each. It is during these discussions that the steering committee should be mindful of the DAT processes and procedures so that the selected data warehousing system is effectively and efficiently integrated within the RTI or PBIS framework. Ultimately, it is of no benefit to students if a wonderfully created data warehousing system is not used or underutilized. If a school district, however, thoughtfully considers its needs and carefully selects a data warehousing system, the decision-making process within RTI and PBIS can be greatly enhanced.

References

Algozzine, B., Horner, R. H., Sugai, G., Barrett, S., Dickey, S. R., Eber, L., & Tobin, T. (2010). *Evaluation blueprint for school-wide positive behavior support*. Eugene: National Technical Assistance Center on Positive Behavior Interventions and Support. www.pbis.org. Accessed 27 Feb 2015.

Anderson, N. A., Childs, K., Kincaid, D., Horner, R. H., George, H., Todd, A., & Spaulding, S. A. (2012). *Benchmarks of advanced tiers v. 2.5*. Unpublished instrument, Educational and Community Supports, University of Oregon, and University of South Florida.

Barnett, D., Daly, E., Jones, K., & Lentz, F. E. (2004). Response to intervention: Empirically-based special service decisions for increasing and decreasing intensity using single case designs. *The Journal of Special Education, 38,* 66–79.

Boneshefski, M. J., & Runge, T. J. (2014). Addressing disproportionate discipline practices within a schoolwide positive behavioral interventions and supports framework: A practical guide for calculating disproportionality rates. *Journal of Positive Behavior Interventions, 16,* 149–158. doi:10.1177/1098300713484064.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12,* 133–149. doi:10.1177/1098300709334798.

Brown-Chidsey, R., & Steege, M. W. (2005). *Response to intervention: Principles and strategies for effective practice*. New York: Guilford Press.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394. doi:10.1177/073428290502300406.

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children, 78,* 356–373.

Christ, T. J., Zopluoglo, C., Monaghen, B. D., & Van Norman, E. R. (2013). Curriculum-based measurement of oral reading: Multi study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology,* 51(1), 19–57. (in press).

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson Prentice Hall.

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program* (2nd ed.). New York: Guilford Press.

Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33,* 429–443.

Eber, L., Phillips, D., Hyde, K., Breen, K., Rose, J., Lewandowski, H., & Upreti, G. (2010). *Illinois positive behavior interventions & supports (PBIS) network: FY10 annual progress report*. Springfield: Illinois PBIS Network. www.istac.net. Accessed 27 Feb 2015.

Educational and Community Supports. (2011). *School-wide information system* [On-line data-base]. Eugene: Author.

Elliott, S. N., & Gresham, F. M. (2007). *Social skills improvement system—Performance screening guide*. San Antonio: Pearson.

Fox, L., Hemmeter, M. L., & Jack, S. (2010). *Early childhood program-wide PBIS benchmarks of quality*. Unpublished instrument, University of South Florida.

Fuchs, L. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice, 18,* 172–186. doi:10.1111/1540-5826.00073.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–500.

Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment, 31,* 16–28. doi:10.1177/0734282912451971.

Good, R. H., Kaminski, R. A., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., & Wallin, J. (2011). *DIBELS Next*. Eugene: Dynamic Measurement Group.

Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention and assessment: A meta-analysis of racial differences. *Journal of Special Education, 37,* 67–81.

Howe, K. B., Scierka, B. J., Gibbons, K. A., & Silberglitt, B. (2003). A school-wide organization system for raising reading achievement using general outcome measures and evidenced-based instruction: One education district's experience. *Assessment for Effective Intervention, 28,* 59–72. doi:10.1177/073724770302800307.

Howell, K. W., & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making* (3rd ed.). Belmont: Wadsworth, Cengage Learning.

Illinois PBIS Network. (2012). *FY12: End of year report*. Springfield: Author. http://www.istac.net. Accessed 27 Feb 2015.

Kamphaus, R. W., & Reynolds, C. R. (2007). *Behavior assessment system for children—Second edition (BASC-2): Behavioral and emotional screening system*. Bloomington: Pearson.

Kelley, B. (2008). Curriculum-based evaluation in math. In A. Thomas & J. Grimes (Ed.), *Best practices* *in school psychology V* (pp. 419–438). Bethesda: National Association of School Psychologists.

Kincaid, D., Childs, K., & George, H. (2005). *School-wide benchmarks of quality*. Unpublished instrument, University of South Florida.

Kovaleski, J. F., & Pedersen, J. (2008). Best practices in data analysis teaming. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology, V* (Vol. 2, pp. 115–129). Bethesda: National Association of School Psychologists.

Kovaleski, J. F., & Pedersen, J. A. (2014). Best practices in data-analysis teaming. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative descision making,* (pp. 99–120). Bethesda: National Association of School Psychologists.

Kovaleski, J. F., VanDerHeyDen, A. M., & Shapiro, E. S. (2013). *The rti approach to evaluating learning disabilities*. New York: Guilford Press.

Marston, D., Muyskens, P., Lau, M., & Canter A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice, 18,* 187–200. doi:10.1111/1540-5826.00074.

National Association of State Directors of Special Education. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria: Author.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read*. Bethesda: Author.

NCS Pearson, Inc. (2011). *AIMSweb*. Austin: Author.

Northwest Evaluation Association. (2004). *Measures of academic progress*. Lake Oswego: Author.

Renaissance Learning. (1996). *STAR Reading*. Wisconsin Rapids: Author.

Reschly, D. J., Tilly, W. D., & Grimes, J. (Eds.). (1999). *Special education in transition: Functional assessment and noncategorical programming*. Longmont: Sopris West.

Runge, T. J., Staszkiewicz, M. J., & O'Donnell, K. H. (2012). *2010–2011 PAPBS network SWPBIS executive summary*. Indiana: Indiana University of Pennsylvania. http://www.papbs.org. Accessed 27 Feb 2015.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford Press.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review, 32,* 317–342.

Skiba, R. J., Horner, R. H., Chung, C., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40,* 85–107.

Spaulding, S. A., Irvin, L. K., Horner, R. H., May, S. L., Emeldi, M., Tobin, T. J., & Sugai, G. (2010). Schoolwide social-behavioral climate, student problem behavior, and related administrative decisions: Empirical patterns from 1,510 schools nationwide. *Journal of Positive Behavior Intervention, 12,* 69–85. doi:10.1177/1098300708329011.

Sprague, J., Colvin, G., & Irvin, L. (2002). *The school safety survey version* 2.0. Unpublished instrument, The Institute on Violence and Destructive Behavior.

Steege, M. W., & Watson, T. S. (2009). *Conducting school-based functional behavioral assessments* (2nd ed.). New York: Guilford Press.

Stewart, L H., & Silberglitt, B. (2008). Best practices in developing academic local norms. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 225–242). Bethesda: National Association of School Psychologists.

Sugai, G., & Horner, R. H. (2009). Schoolwide positive behavior support. In W. Sailor, G. Sugai, G. Dunlap, & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 307–326). New York: Springer.

Sugai, G., Lewis-Palmer, T., Todd, A. W., & Horner, R. H. (2005). *School-wide evaluation tool version 2.1*. Eugene: Educational and Community Supports.

Sugai, G., Horner, R. H., & Lewis-Palmer, T. (2009a). *Effective behavior support: Team implementation checklist version 3.0*. Eugene: Educational and Community Supports.

Sugai, G., Horner, R. H., & Todd, A. (2009b). *Effective behavior support: Self-assessment survey version 3.0*. Eugene: Educational and Community Supports.

Tilly, W. D., III. (2008). The evolution of school psychology to science-based practice: Problem-solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology,* V (Vol. 1, pp. 17–36). Bethesda: National Association of School Psychologists.

Todd, A. W., Horner, R. H., Newton, J. S., Algozzine, R. F., Algozzine, K. M., & Frank, J. L. (2011). Effects of team-initiated problem solving on decision making by schoolwide behavior support teams. *Journal of Applied School Psychology, 27,* 42–59. doi:10.1080/15377903.2011.540510.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40,* 7–26. doi:10.1016/S0022-4405(01)00092-9.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). The development and validation of a process for screening referrals to special education. *School Psychology Review, 32,* 204–227.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential pitfalls. *Learning Disabilities Research & Practice, 18,* 137–146.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Multi-Tiered Systems of Support and Evidence-Based Practices

Karen C. Stoiber and Maribeth Gettinger

Inherent to a multi-tiered service delivery model is the provision of evidence-based, high-quality instruction and intervention to all students. EBPs refer to effective, research-based strategies and programs, including supplemental differentiated support, shown to produce positive outcomes (Forman et al. 2013; Novins et al. 2013). The term "evidence based" has been described in the educational and psychological literature as the level of evidence that supports the efficacy, generality, and use of a practice as indicated by research (Stoiber and DeSmet 2010). In general, the more specific term "evidence-based interventions," or EBIs, refers to prevention, intervention, or treatment programs having strong scientific support or research evidence (e.g., at least one published study using strong design features and demonstrated positive measured outcomes; Kratochwill and Stoiber 2002). EBPs are defined as practices that integrate the best available research with clinical expertise in the context of student characteristics, culture, and preferences. The term EBP is distinguished from the term EBI, in that an EBP may be based on (a) demonstrated research-based outcomes and/or (b) context-specific, data-based decision-making that incorporates data collected by the practitioner for progress monitoring or program evaluation purposes (Stoiber and DeSmet). At present, the term EBP is being used more frequently in the literature and corresponds to the language in current school reform policies. EBPs constitute an essential feature of successful implementation of MTSS.

Although a wide range of effective preventive and intervention EBPs has been developed for application in clinical settings, especially for addressing child and adolescent mental health, fewer have been evaluated within school-based settings. For example, Novins et al. (2013) conducted a review of empirical studies examining EBPs to improve mental health outcomes in children and adolescents. Just over one-third of the studies (36 % or 16 studies) identified by Novins et al. were determined to use a true experimental design such as including randomized control groups, with observational and descriptive studies being most common. Of the 16 studies found to meet their criteria for methodological rigor and relevance, only 25 % (4 of 16) were conducted in school settings. Interestingly, the majority of EBP studies meeting criteria, including school-based investigations reviewed by Novins et al., targeted substance abuse. Miller et al. (2009) performed a review of studies examining suicide prevention programs based on the *Evidence-Based Interventions in School Psychology Procedural and Coding Manual* (Kratochwill and Stoiber 2002). Their review found that only 2 of 14 school-based studies demonstrated significant positive

K. C. Stoiber (✉)
Univerisity of Wisconsin, Milwaukee, WI 53211, USA
e-mail: Kstoiber@uwm.edu

M. Gettinger
University of Wisconsin, Madison, WI 53706, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_9

statistical effects on the primary measures. In addition, less than a quarter (23 %) of the studies identified the intervention components linked to the primary outcomes, and similarly few studies (23 %) provided information regarding program implementation integrity. Miller et al. reported only one suicide prevention study (7.6 %) provided promising evidence of educational/clinical significance. Clearly, more high-quality research is needed aimed at the development of EBPs, especially for addressing students' targeted mental health concerns.

Despite the need for continued development of EBPs, there are a number of empirically supported programs and strategies appropriate for application in school settings. Several groups have established criteria for reviewing the effectiveness of prevention and intervention programs and practices such as the *Promising Practices Network* (http://www.promisingpractices.net) and the federally funded *What Works Clearinghouse* 2007 (WWC; http://www.whatworks.ed.gov). For example, the *Promising Practices Network* offers descriptions of interventions aimed at improving academic difficulties (e.g., peer-assisted learning strategies; PALS) and for addressing social–emotional and behavioral concerns (e.g., Social Decision-Making/Problem-Solving; Good Behavior Game). In addition, the website offers an overview of specific academic interventions (e.g., Reading Recovery) and social–emotional–behavioral programs (e.g., Second Step violence prevention). The WWC reviews programs for school-based implementation and provides evidence ratings on the level of research to support practices for a variety of concerns (e.g., dropout prevention, reducing behavior problems in the elementary school classroom, improving adolescent literacy, effective literacy, and English language instruction for English learners). Thus, although an analysis of the reasons for the limited uptake of EBPs in schools suggests a need for continued focus on the development of EBPs, it also points to limited EBP dissemination and implementation as culprits in the research-to-practice gap. In this regard, an effort to increase the use of EBPs by school professionals would seem to be enhanced by increasing an understanding of which strategies have been shown to be effective as well as attention to considerations in their implementation. A key challenge for schools in facilitating optimal and effective implementation of MTSS is ensuring that school professionals have acquired knowledge and skills in EBPs.

Implementation of EBPs for Successful MTSS A critical step in helping schools and educators be successful in EBP implementation within MTSS is to foster knowledge of what strategies or approaches work to address particular academic and social behavioral problems. That is, education professionals are more prone to accept, support, and implement EBPs when they have an understanding not only of the "what" but also have skills regarding the "when and how" of prevention and intervention strategies (Forman et al. 2013). Further, the research-practice gap should lessen when school-based practitioners view EBPs as feasible and readily incorporated into what they do on a daily basis. Thus, for EBPs to move from being expected to becoming commonplace in US schools, it is essential to target school psychologists, teachers, and other school-based professionals' EBP-related knowledge, skills, and attitudes.

Several prominent researchers have documented educators' limited knowledge of EBPs for supporting students with academic and mental health issues. Stormont et al. (2011) examined teachers' knowledge of ten EBIs and resources to support children with social–emotional and behavioral difficulties. These ten programs had met criteria for empirical support (see Blueprint's for Violence Prevention; WWC), and included Positive Behavioral Interventions and Supports (PBIS), Promoting Alternative Thinking Strategies (PATHS), First Step to Success, Olweus' Bully Prevention, Positive Parenting Program (Triple P), Second Step, Good Behavior Game, Coping Power, Incredible Years, and Coping Cat. Of the ten possible evidence-based programs shown to be empirically supported, the majority of teachers in their study recognized only one EBP (i.e., PBIS). For the other nine programs, 10 % or less of the teachers endorsed the programs as evidence based. Further, more than half of the educators were uncertain whether functional assessment and intervention planning occurred at their school, and more than 75% were unclear whether any data were collected at their school on the effectiveness of school-based mental health programs. These results are similar to Stoiber and Vanderwood's (2008) study of preferred practices and reported levels of competence in urban school psychologists. School psychologists in their study reported both limited experience and low competence in practices such as conducting functional assessments and monitoring academic and behavioral interventions despite endorsing them as among the most important practices.

Evidence-Based Practices for Improving Academic and Social Behavioral Outcomes

The following sections review the current research knowledge of EBPs for implementation of MTSS to respond to students' academic and social–emotional–behavioral needs. For MTSS to be successful, students' academic and social behavior concerns should be addressed in a comprehensive and cohesive manner. Nonetheless, it is important to recognize that research-based knowledge differs depending on the instructional content and area of concern. More specifically, the empirical base regarding effective instruction and intervention in academic content domains is generally considered better developed than in social–emotional and behavioral domains. Well-regarded syntheses of research on effective literacy and mathematics instructional practices are available (see National Early Literacy Panel 2009 for a synthesis of literacy instruction; and Slavin and Lake 2008 and Slavin et al. 2009 for a synthesis of mathematics instruction). In addition, national research centers have documented EBPs that foster and support the development of reading (e.g., Center for Early Literacy Learning; National Institute for Literacy; WWC). Collectively, these evidence sources comprise a scientific foundation for schools to teach reading and mathematics effectively.

In contrast, empirical knowledge of what works to prevent challenging behavior and promote resilience is less developed, and hence, less available and applied in the schools. In particular, an understanding of how to counter student difficulties, while at the same time improve children's social competence, is needed. In a recent study, Stoiber and Gettinger (2011) demonstrated that both targets for change (that is, reducing challenging behavior *and* improving social competencies) can be viable outcomes with children at risk for developing challenging behavior, when applied using a comprehensive, systematic intervention framework. Nonetheless, the research literature indicates that different social–behavioral intervention techniques demonstrate varying levels of efficacy in promoting resilience and/or treating problematic social–emotional–behavioral patterns in children (Doll et al. 2005; Langley et al. 2010; Vaughn et al. 2009). One reason for the uneven success of social–behavioral treatments is that many of the procedures undertaken in schools have been less comprehensive, integrative, and systematic than needed. Further, there is evidence that educators and other school professionals may be reluctant to use procedures shown to be effective for addressing and improving social–emotional and behavior concerns, such as conducting functional assessments and monitoring student progress toward expected outcomes, as they lack knowledge in these procedures or may view them as complex and time-consuming (Gettinger and Stoiber 2006; Gresham et al. 2013; Stormont et al. 2011).

Recently, some applications of an integrated MTSS model have begun to emerge. For example, McIntosh and colleagues (McIntosh et al. 2006, 2010), Lane and associates (Lane et al. 2009, 2012), and Stoiber (in press) have described conceptual frameworks for the integration of both academic and behavior support within a multi-tiered system. These frameworks emphasize the importance of optimizing system-level and organizational support through a combined focus on academic and social–behavioral performance indicators. In that, an MTSS approach for both reading and behavior provides support for all students through implementation of EBP at all tiers, there also have been preliminary efforts to examine the feasibility of integrated MTSS models. One example is the Michigan Integrated Behavior and Learning Support Initiative (MiBLSi), which is funded through the Michigan Department of Education with the goal of improving both behavior and reading skills in schools statewide (Ervin et al. 2006).

The MiBLSi initiative grew out of an understanding of the important and well-documented linkage between students' behavior and academic achievement. The use of an integrated MTSS approach allows literacy and behavioral intervention components to positively influence each other. That is, as reading improves, students are less likely to engage in disruptive behaviors. Likewise, as instructional time increases due to less time spent addressing problem behaviors, so does reading achievement. Since the initial implementation of MiBLSi in 2004, the percentage of students statewide meeting reading benchmarks has increased, on average, 5% every year, and the rate of office disciplinary referrals has decreased, on average, 10% per year (Hartsell and Hayes 2012). Moreover, participating schools have experienced, on average, a 21% reduction in special education referrals and 26% drop in identification rates, particularly between the first and second years of implementation (Michigan Department of Education 2012).

Because schools are just beginning to apply MTSS as a means to organize and deliver EBPs for students who are at risk for reading and literacy difficulties as well as those with social–emotional and behavioral concerns, much of the knowledge on EBPs is content specific. That is, for the most part, there is a knowledge base linked to EBPs for improving academic outcomes, and a separate knowledge base on EBPs for promoting social–emotional and behavioral competence. Nonetheless, there are several strategies for implementation of MTSS across domains that are considered evidence based, which are summarized in Table 1. These practices are viewed as corresponding to a comprehensive school improvement effort for improving students' academic and social–emotional performance outcomes (Lane et al. 2012; McIntosh et al. 2010).

Table 1 Implications for practice: Key characteristics of evidence-based and differentiated instruction

Explicit and/or intentional instruction occurs regularly whereby key concepts and learner expectations are taught *purposefully*
Key concepts, competencies, and learner outcomes for academic (e.g., vocabulary, listening comprehension, mathematical problem-solving) and social–behavioral (e.g., being respectful) competence development are *targeted and modeled* across all multi-tiered levels
Preventative interventions occur early and are based on relevant screening and assessment so as to maximize attempts to *close the gap* between at-risk and typically developing students
Instructional approaches at all tiers promote *active involvement* and *opportunities to respond* along with a variety of appropriate ways to be engaged and reengaged
Instruction and intervention approaches *maximize student engagement and motivation* by including clear expectations for performance balanced with flexibility and the use of choice
Keystone behaviors that promote both academic and social–behavioral success (e.g., demonstrate responsible behavior and self-regulation) are emphasized across learning environments
Instruction *progresses logically* within content domains and moves from easier concepts/skills to more challenging ones
Opportunities to practice newly learned concepts and skills are provided with varied degrees of teacher support (e.g., incorporating teacher- and peer-mediated strategies and independent practice) and different learning contexts to maximize student learning and motivation
Mastery of expected academic and social–behavioral outcomes is *monitored carefully* so that reteaching, instructional modeling, and corrective feedback occur as needed
Provision of diverse and varied *small-group instruction* wherein students are grouped based on a variety of indicators (particular skill development, need, interest)
Resources: Denton 2012; Gettinger and Stoiber 2012; Stoiber and Gettinger 2012; Stoiber in press; Vaughn and Chard 2006

Because the knowledge base on EBPs has generally been domain specific (derived through investigations focused on either the academic *or* social–emotional functioning), applications of MTSS and EBP in academic and social domains are reviewed below in separate sections. This content-specific knowledge of EBPs is viewed as foundational for helping districts and schools increase implementation of MTSS that aims to promote both academic and behavioral success among students.

Applications of MTSS and EBP in Academic Domains

Increasingly, schools are implementing tiered systems of support in an attempt to meet students' diverse literacy needs and prevent the emergence of academic difficulties (Fletcher and Vaughn 2009; Fuchs and Vaughn 2012). Although multi-leveled instruction has been promoted as EBP for all students, it has particularly been advocated for those students who struggle with reading in the primary and later elementary grades (Gersten et al. 2009). There are a variety of models associated with MTSS and EBP for literacy development, nonetheless, the common components across all applications include (a) consistent, high-quality evidence-based core instruction for all learners, (b) screening and progress-monitoring procedures to predict responsiveness to tiered instruction, and (c) more intensive interventions designed to supplement classroom instruction for students identified as being at risk based on screening indices or progress-monitoring measures (Kovelski and Black 2010; O'Connor and Freeman 2012). Typically, tiered models consist of three levels of differentiated instructional intervention (Lane et al. 2012). The tier levels are often referred to as primary or universal (tier 1), secondary or targeted (tier 2), and tertiary or intensive (tier 3).

Burns et al. (2005) synthesized the research related to multi-tiered instruction in academics, predominantly reading and literacy. These researchers found positive effects for both system-level outcomes (special education referral, grade retention, and time spent in special education) and individual student-level outcomes (achievement, growth estimates, and academic engaged time) in sites implementing tiered academic instruction. Overall, implementation of multi-tiered models increased student outcomes whether implementation was school based (mean effect size = 0.94) or researcher implemented (mean effect size = 1.14). Additional studies of multi-tiered literacy models have provided further evidence of (a) higher reading outcomes for all students over time, (b) a decrease in the number of at-risk students identified for tier 2 instruction, (c) accelerated learning (higher slope of progress) among students receiving tier 2 and tier 3 instruction, (d) a decline in the number of students in special education, and (e) a reduction in disproportionate placement of students who are male, minorities, or English language learners (ELL; Gettinger and Stoiber 2012; Torgesen 2009; VanDerHeyden et al. 2007; Vaughn et al. 2009; Wanzek and Vaughn 2011). EBPs associated with each of the tiers in a multi-tiered model are described next. To date, literacy has been the primary focus of multi-tiered delivery models. Thus, most of the knowledge of EBPs in implementing MTSS relates to the area of literacy, and is the primary area highlighted.

Universal or Tier 1 Interventions Within the first tier, all students receive instruction using a comprehensive, evidence-based core program (often aligned with state standards) provided by the general education teacher. A well-implemented and effective core curriculum program at tier 1 should help ensure that students have had adequate opportunities to learn critical content, and thus lead to fewer students requiring intervention (Vaughn et al. 2009). Teachers, nonetheless, will typically need to make adjustments at tier 1, even with an effective core curriculum. These instructional adjustments within the first tier will likely require adapting or supplementing the core curriculum to meet diverse student learner needs (Fuchs and Vaughn 2012). Examples of differentiated reading instruction at tier 1 include deciding whether a student will benefit from an additional focus on code-related (e.g., letter identification, letter–sound recognition) or meaning-based instruction (e.g., vocabulary, reading comprehension). Not only do teachers at tier 1 need to be able to design such instructional differentiation but they also need to have knowledge regarding instructional pedagogy such as determining when and how to alter the type and intensity of instruction and which methods should be applied to scaffold student learning (e.g., peer tutoring, peer coaching, small group instruction). To facilitate classroom teacher decision-making at the first tier, the use of a variety of screening, progress monitoring, and assessment indices is suggested (VanDerHeyden et al. 2007). Example assessment approaches include informal inventories of letter–sound knowledge or sight words, oral reading fluency indices, curriculum-based measures of vocabulary and reading, and indicators of reading comprehension (Denton 2012; Gettinger and Stoiber 2012).

The intent of tier 1 in reading is to deliver high-quality instruction that has been shown to promote key literacy outcomes. Effective tier 1 instruction is well organized and incorporates planned lessons that target key literacy outcomes including phonemic awareness, phonics, word recognition, fluency, vocabulary, and comprehension. There is a well-established evidence base documenting improved literacy outcomes when teachers implement validated practices that focus on these targeted reading skills within tier 1 (Fuchs and Vaughn 2012; Foorman 2003; Justice 2006; Torgesen 2009). At the early grade levels, an emphasis on vocabulary should include development of background knowledge in a variety of teaching contexts, automatic recognition of high-frequency irregular words, and ample opportunities for students to learn through such methods as repeated reading (Denton 2012; Gersten et al. 2009). Although less is known about effective tier 1 instruction when serving nonnative English speakers, a similar focus on key components (e.g., explicit instruction in phonemic awareness, oral reading fluency, vocabulary development, etc.) is supported for ELLs in tier 1 (Vanderwood and Nam 2008). In addition, ELLs benefit from instruction focused on their specific needs in oral language development, such as extended opportunities to learn and practice vocabulary, and to use newly learned words in listening and speaking as well as in reading and writing (Crosson and Lesaux 2010). The focus on oral language for ELLs is further supported by Crosson and Lesaux's finding that the relationship between reading fluency and comprehension in native English readers appears to be moderated by ELLs' oral language development.

High-quality tier 1 instruction should lead to fewer students needing additional support and, in theory, enables 75–80 % of students to achieve expected literacy benchmarks. Thus, even with effective and differentiated tier 1 instruction, as high as 20–25 % of students will fail to develop proficiency in reading skills (Fuchs and Vaughn 2012). For some students, tier 1 instruction moves at too rapid a pace, provides insufficient practice or opportunities to respond, or does not focus on skills with sufficient intensity or duration (Stanovich and Stanovich 2003; Fuchs and Vaughn 2012). When tier 1 instruction is not adequate, students are provided more explicit and intentional instruction within higher tiers. Prior to moving students into higher-tiered interventions, however, it is important to determine whether the tier 1 instruction was sufficient. An examination of the adequacy of tier 1 should especially occur when schools or districts witness more students than the expected 20 or 25 % as failing to meet established performance benchmarks. To begin to examine whether tier 1 is sufficient in promoting expected student outcomes, educators can explore key questions that underlie the basis of MTSS. Table 2 presents ten orienting questions to facilitate such an analysis of the foundational characteristics of MTSS, and, more specifically, the quality of instruction occurring at tier 1.

The second and third instructional tiers involve evidence-based programs and practices designed to reinforce and supplement the core reading program. Within an MTSS approach, students whose benchmark screening data indicate some, but not high, risk for reading failure receive tier 2 instruction; students who are at high risk and/or not responsive to tier 2 strategies receive tier 3 instruction. Tier 2 involves instructional programs aimed at a level of skill development further along a continuum of skill acquisition than what is targeted by tier 3 instruction, which typically targets more basic or foundational skills. In practice, tier 2 instruction begins as soon as possible after students have been identified as falling below grade-level expectations (assuming that tier 1 was deemed sufficient) and is usually implemented between 6 and 10 weeks.

Targeted or Tier 2 Interventions Instructional interventions provided at the higher tiers should not replace the core curricula, but rather aim to enhance and supplement students' learning. Meta-analytic and descriptive research reviews provide strong evidence for the effectiveness of supplemental, tier 2 instruction (Elbaum et al. 2000; Gersten et al. 2009). Tier 2 interventions may be delivered following a standard protocol for instructional interventions that permit increased practice opportunities for skill development (e.g., reading, mathematics). In the area of literacy and mathematics, researchers have established support for a number of key supplemental strategies for preschool through secondary students, including more targeted instruction with explicit modeling and teaching, greater attention to daily review activities, increased teaching for generalization and guided instruction, and multiple opportunities for practice including independent practice (Denton 2012; Fuchs and Vaughn 2012; Jones et al. 2012). Thus, in addition to covering content yoked to the core curriculum, tier 2 instruction should incorporate more exposure, more time, and more opportunities to learn. The practice of "supplementing more, more, and still more" should be apparent at tier 2.

Based on evidence presented by WWC (http://www.whatworks.ed.gov) and Gersten et al. (2009), to be effective, tier 2 reading instruction should target literacy skills for which students require additional support, occur three to five times weekly (20–40 min each session), and incorporate frequent opportunities to practice skills with teacher feedback. Several researchers have evaluated the benefits of tier 2 instruction when provided very early in the primary grades (e.g., kindergarten) versus later (winter of first grade). Example tier 2 strategies in the early grades include using repeated readings and a focus on phonemic awareness. There is some evidence that kinder-

Table 2 Ten orienting questions for examining foundational characteristics of RTI

1. Is there an effective core curriculum matched to state/district/school/center-based learner goals and expectations?
2. Is there a coordinated and aligned system of effective positive behavior support that includes social competence support/instruction and prevention and intervention strategies?
3. Are there well-articulated and meaningful goals, objectives, and/or benchmarks representing academic and social–behavioral domains and are they being used to structure instruction?
4. Is there a brief, repeatable, formative assessment of progress toward benchmarks or important learner goals that is sensitive to intervention?
5. Is there a flexible delivery of instruction that provides sufficient opportunities for practice and learning in a variety of instructional contexts (e.g., whole group instruction, small-group instruction, peer-assisted learning, independent practice)?
6. Do teaching staff use differentiated instructional strategies, including small-group instruction and repeated practice opportunities which provide effective intervention to students at risk to prevent more severe difficulties?
7. Are there decision rules or procedures for staff to mobilize intensive prevention resources very early, before serious learning difficulty and/or social–behavioral problem behavior occur?
8. Is there a school-wide collaborative process to coordinate resources within the school/district/community context to accomplish tiered prevention and intervention efforts?
9. Have teaching staff received training and practice, and do they use a broad range of "evidence-based" prevention strategies, teaching strategies, and alternative response strategies to foster positive learning and behavior in all children and to address their diverse needs?
10. Is there a mechanism or structure for the provision of consultation and/or coaching to ensure that instruction/intervention at all levels is high quality, delivered with fidelity, and evaluated to be consistent with evidence- and/or empirically validated processes and programs?

RTI response to intervention

garten may be a critical "window of opportunity" in the prevention of reading difficulties. That is, the provision of tier 2 instruction in kindergarten may be particularly potent in preventing the need for more intense intervention at a later time (Denton 2012). Other researchers, however, have noted that kindergartners who had tier 2 benefited from continued monitoring of skills during the primary grades, and some students required subsequent intervention (Kamps et al. 2008). More research is needed to determine optimal timing of tier 2 for maximizing students' reading success, especially within the primary grades.

At the secondary level, the focus of tier 2 instruction shifts to remediation and content-specific recovery. In addition to the short-term outcome of helping students pass core courses/exams, it has the long-term goal of promoting their graduation (Pyle and Vaughn 2012). Several studies have suggested that intervention emphasizing comprehension (Graves et al. 2011) and vocabulary instruction yields stronger effects for struggling students at the secondary level compared to their elementary-level counterparts (Coyne et al. 2010). Coyne et al. have documented that effective vocabulary learning for at-risk adolescents incorporates motivational strategies coupled with more complex word study (e.g., multisyllabic, vocabulary words linked to subject content). Their supplemental tier 2 vocabulary intervention targeted key vocabulary words through the provision of multiple opportunities for practice and immediate feedback.

There also is support for a multicomponent reading intervention that targets several skill areas (e.g., phonics, vocabulary, oral fluency, reading comprehension) as potentially beneficial with upper-elementary-, middle-school-, and high-school-aged students who struggle in reading (Canter et al. 2008; Graves et al. 2011). Graves et al. designed a tier 2 intervention for sixth-grade students, which consisted of a multiple evidence-based programs to target several key skills, word analysis, fluency building, comprehension, and vocabulary. That is, the tier 2 intervention consisted of different reading programs to address word analysis (i.e., Corrective Reading, REWARDS), fluency (i.e., Read Naturally) and vocabulary and comprehension (i.e., Daybook for Critical Reading and Writing). They found that intensive tier 2 instruction was beneficial for improving middle school students' oral reading fluency and reading comprehension, but no significant effects occurred for vocabulary or on a maze syntactic sentence completion measure. The results of studies such as those by Graves et al. point to the need for more attention to designing and refining targeted interventions for older struggling students. Though the knowledge base on tier 2 interventions with ELLs is somewhat limited, support exists for continued explicit focus on their oral language skills at the secondary level, including vocabulary development and providing extended opportunities for practicing learned words in communicating and in listening activities (Denton 2012; Graves et al. 2011; Vaughn et al. 2011).

Intensive or Tier 3 Interventions At the end of the designated intervention period, students may discontinue tier 2, receive another round of tier 2 instruction, or move to more intensive instruction in tier 3 (Fuchs and Vaughn 2012). The third tier consists of instruction that is customized for students who continue to struggle despite having received evidence-based universal and supplemental reading instruction (approximately 5% of students). Research has documented several effective interventions for students with severe reading difficulties that inform tier 3 reading instruction (Burns et al. 2008; Fletcher and Vaughn 2009; Kamps et al. 2008). Typically, intensive tier 3 interventions are differentiated from lower tiers by several distinguishing characteristics. These characteristics include requiring more time and resources, being implemented by an interventionist rather than the classroom teacher, and conducting progress monitoring more frequently to facilitate improved responding and instruction appropriately matched to the student's skills.

At the secondary level, there also exist some specific EBPs in MTSS. For example, these students need strategy instruction to decode and break multisyllabic words into word parts and explicit instruction in reading comprehension skill.

Due to the more extensive length of time during which students have experienced achievement gaps, secondary level students may require multiple years of intensive intervention and remediation to reach expected levels (Pyle and Vaughn 2012). Thus, the parameters used to determine the duration and level of intensity of tiered interventions will likely need to be fundamentally different for students at more advanced grade levels.

Applications of MTSS and EBP in Social–Emotional and Behavioral Domains

Similar to developments in reading and literacy, there has been increased attention to the use of scientifically based behavioral interventions and multiple levels of support to prevent the development of problem behaviors and address the needs of students with behavior challenges. Multi-tiered models for the development of social–emotional and behavioral competence incorporate a continuum of behavior support comprising intervention levels designed to prevent, respond to, and/or reduce challenging behaviors (Iovannone et al. 2009; Stewart et al. 2007). Consistent with MTSS in reading and other academic areas, this continuum is typically conceptualized as a three-tiered approach, with the intensity of intervention matched to student needs. Because a key component of social–behavioral MTSS applications is to foster positive student outcomes, it naturally shifts the blame from child-centered to ecological-focused factors (Gresham et al. 2013; Stoiber and Gettinger 2011). The majority of research-based MTSS applications for behavioral concerns are conceptualized and implemented within a Positive Behavior Support (PBS) or School-Wide Positive Behavior Support (SWPBS) framework (Horner et al. 2010).

Similar to evaluations of MTSS applications for literacy, studies of SWPBS tend to focus on the effectiveness of interventions within separate intervention tiers. In general, there is empirical evidence that when interventions within each tier (especially tier 1) are implemented with fidelity, there are improved social–behavioral as well as academic outcomes. Multiple studies have demonstrated the positive impact of SWPBS on reducing suspensions and office discipline referrals (ODRs), promoting school safety, increasing prosocial behavior while decreasing problem behavior, and enhancing student achievement outcomes (Bradshaw et al. 2008, 2010; Horner et al. 2009; Lassen et al. 2006).

Horner et al. (2010) conducted a review of 46 studies published since 2000 and concluded that SWPBS has sufficient evidence to warrant large-scale implementation. In their review of SWPBS research, Horner et al. applied a set of well-established criteria for determining whether a practice is evidence based (e.g., procedures are defined with operational precision, research employs valid and reliable measurement). Two studies, in particular, met all criteria including the use of a randomized control research design. The first of these studies randomly assigned 21 elementary schools to implement SWPBS and 16 schools to a control group. Data were collected over a 5-year period in all schools and revealed a reduction in student suspensions, fewer ODRs, and improved academic achievement for SWPBS schools (Bradshaw et al. 2010). In the second study, researchers used a randomized, wait-list controlled trial to assess the effect of SWPBS on student outcomes in elementary schools in Hawaii and Illinois (Horner et al. 2009). The use of SWPBS was related to improvements in the perceived safety of the school setting, low ODRs, and a significant increase in the proportion of third graders meeting or exceeding state reading assessment standards. Substantial evidence validates the use of PBIS/SWPBS as an effective MTSS approach to prevent challenging behaviors, to close the gap between identification and intervention, and promote success for all learners (Chitiyo et al. 2012; Horner et al. 2010; Lassen et al. 2006; Sailor et al. 2009). There is less agreement, however, regarding how often and which specific strategies/approaches used in tier 2 or 3 differ from those in tier 1 for behavioral (compared to academic) applications of multi-tiered approaches (Hammond et al. 2013; Lindstrom 2013). Thus, strategies or approaches presented below under tier 1 may be used with a more intensive and customized manner at upper tiers. Similarly, strategies and programs described in tier 2 may be used in tier 3 and vice versa. Specific decisions regarding the amount and type of intervention used will depend on the presenting concern or possible diagnostic category being served (e.g., autism, attention deficit hyperactivity disorder; ADHD; see Hammond et al. 2013; Lindstrom 2013) and on how a state, district, and/or school have conceptualized multi-tiered services.

Universal or Tier 1 Interventions Within the MTSS framework, EBIs are organized into a tiered continuum that, first, provides all students with a positive classroom environment and appropriate behavior support (tier 1), and then sequences an array of interventions of increasing intensity to accommodate students whose behaviors are not responsive to tier 1 support (Sailor et al. 2009). Although there is flexibility within the MTSS framework for customizing PBS for individual schools and districts, certain practices are standard across all school-based applications. Specifically, in the first intervention tier, a small number (three to five) of positively stated, operationalized behavioral expectations (e.g., be respectful, be safe, be responsible) are taught to all students using explicit and systematic instructional procedures. It is recommended that each school-wide expectation is posted throughout the school, including classrooms and common areas (e.g., corridors, lunch room, gym). In addition, students receive frequent recognition and positive consequences for meeting expectations, and a continuum of logical consequences for clearly defined unacceptable behavior is explained and administered. For SWPBS to be effective, it requires buy-in from 80% of school staff for a 2-year period (Horner and Sugai 2009). Classroom teachers should create a positive classroom environment and be able to employ a range of consequences or intervention strategies for addressing problem behavior and deliver them consistently.

Recent indicators suggest that many children come to school with limited social–emotional competencies and would benefit from a caring, encouraging environment aimed at enhancing their motivation and sense of belonging (Peterson et al. 2013). As the focus of MTSS is on positive support, a social–emotional learning (SEL) approach is well aligned with the preventative intent of the foundational tier. The SEL approach stems from a "strengths-based" approach to student behavior with an emphasis on ecological perspectives in favor of targeting child-focused deficit factors is inherent in the SEL philosophy. The primary aims of SEL are to achieve five interrelated positive social competencies in students, self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (Collaborative for Academic, Social, and Emotional Learning, CASEL; 2005). The application of SEL concepts to the school context should lead to students being better adjusted and able to focus on their academic skills, which, in turn, should lead to a reduction in social–emotional distress and conduct problems along with improved test performance and grades.

Typically, SEL school-based programs involve the delivery of classroom curricula that incorporate two key sets of educational strategies. The first strategic component includes instructional features whereby SEL skills "may be taught, modeled, practiced, and applied to diverse situations so that students use them as part of their daily repertoire of behaviors" (Durlak et al. 2011, p. 406). Delivery of this SEL program component should be conducted in a developmentally and culturally appropriate manner that fosters health-promoting outcomes and good citizenship. The SEL curricula may also aim to deter specific types of problems, such as bullying, violence, aggression, substance use, and dropping out of school. The second characteristic of SEL approaches is that they promote students' sense of school safety and connection through the provision of responsive teaching and classroom management along with community-building activities across the school environment. Several evidence-based SEL programs are available for use either as a foundational curriculum or to supplement classroom programing with suggested applications at the school-wide level (see for example, selected programs listed on www.

casel.org or the www.ctclearinghouse.com—PATHS; Second Step; Strong Kids/Strong Start/Strong Teens; Social Decision-Making/Problem-Solving Program).

For SEL programs to be effective, it is recommended that they incorporate four essential features represented in the acronym SAFE: (a) follow a *sequenced,* step-by-step approach, (b) incorporate *active* and interactive training, (c) are *focused* on specific goals with sufficient time to address them, and (d) use *explicit* teaching strategies that clarify and support learning expectations (Bond and Hauf 2004; Durlak et al. 2011). A meta-analysis of school-based universal programs was conducted by Durlak et al. to examine the effects on students' development of social competencies and related expected outcomes (i.e., enhanced social–emotional skills, positive attitudes, and positive social behaviors; reduced conduct problems and emotional distress; improved academic performance). Durlak and his associates also explored teacher effectiveness in administering SEL programs and whether multi- or single- (classroom only) component programs were more effective. In addition, they hypothesized that program outcomes would be moderated by use of the four recommended practices (i.e., sequenced, active, focused, and explicit) and by reported program implementation problems.

The meta-analysis by Durlak et al. (2011) was based on 213 studies. More than half of reviewed investigations (56 %) were conducted at the elementary school level, and 47 % employed a randomized design. The majority of SEL programs were classroom based and taught by teachers (53 %); most occurred within urban school contexts (47 %); and a minority (26 %) were multi-component programs. Meta-analysis results supported SEL programs as producing positive effects on student social–emotional competencies as well as on their attitudes toward self, others, and school. The SEL programs also were found to enhance students' behavioral adjustment (i.e., increased prosocial behavior and decreased conduct and internalizing problems).

One of the most important indications stemming from the meta-analysis is that the SEL programming led to improved academic performance, with their results showing that systematic social–emotional curricula boosted student achievement, on average, by 11 percentile points. Another noteworthy finding of Durlak et al. (2011) was that classroom teachers and other school personnel were effective in implementing SEL programs, suggesting that preventative tier 1 programs can be conducted without outside personnel. Further, SEL programs were successfully implemented across educational levels (elementary, middle, high school) and community settings (urban, suburban, rural), though they are studied less frequently in high school and rural settings. Finally, as predicted, the SAFE practices and implementation problems had an impact on student outcomes, pointing to the positive value of well-designed and well-executed programs. Interestingly, the meta-analysis did not demonstrate that multicomponent approaches such as those incorporating parent or school-wide program features produced additional benefits. This finding may be due to a restricted sample as few SEL programs added coordinated school-wide and parent components to the classroom-based programming. Thus, more research examining whether, which, and how additional components might enhance core SEL programing is needed. In addition to manualized SEL interventions, other classroom-level interventions have been shown to effectively promote cooperation and are suggested for use at tier 1. Two examples are the strategic programs, the Good Behavior Game and Red Light/Green Light, which aim to decrease classroom rule violations, while simultaneously creating a positive learning environment (Stoiber 2004).

Intervention strategies that focus more generally on promoting student engagement through environmental support and effective classroom management strategies have been shown to facilitate a positive classroom environment. Ysseldyke and Christenson (2002) identified several key conditions to support learning, such as instructional match, relevant practice, adaptive instruction, and informed feedback, that should be in place at tier 1 to assure high-quality instruction. It is recommended that specific quality indicators are examined by school administrators' routinely conducting a "walk through" of tier 1 practices. Regardless of whether a school adopts a structured set of guidelines or a manualized SEL program, systematic and responsive teaching of appropriate behavior to all students should be apparent at the universal or primary instructional level.

Targeted or Tier 2 Interventions Tier 2 strategies are provided for students who require more structured behavioral interventions, more frequent and contingent behavior feedback, and/or more active supervision and monitoring by adults. At tier 2, social–behavioral programs typically direct greater attention on teaching schoolwide behavioral expectations to at-risk students in small groups. Teachers continue to focus on problem prevention through the provision of frequent recognition and positive consequences to students for meeting expectations. Students in tier 2 also receive systematic teaching of social–emotional skills by building in ample opportunities for them to practice competencies such as engaging peers appropriately, taking turns when talking, giving compliments, or using strategies such as "stop and think" to resist impulsive reactions or criticism. Several researchers (Gettinger and Stoiber 2006; Lane et al. 2012; Sailor et al. 2009; Stoiber and Gettinger 2011) have suggested that to determine the focus of interventions, it is useful to consider high-priority behavioral concerns and, when feasible, to integrate them within academically focused activities. These concerns are then linked to replacement behaviors or goals and subgoals that may be targeted for improvement in higher tiers, either in small groups or on an individual basis.

Several resources offering possible social competence goals to target in a tiered intervention model are available (see Durlak et al. 2011; Stoiber 2004). In designing the intervention, functional assessments can be useful for selecting appropriate social-behavior goals (Gettinger and Stoiber 2006; Gresham et al. 2013; Jones and Wickstrom 2010; Stoiber and Gettinger 2011). As school personnel choose appropriate goals, they should focus on changing behavior that students are capable of learning, keystone competencies (Gettinger and Stoiber 2006) that likely have powerful effects on adjustment, or "access" behaviors that allow entry to beneficial environments (e.g., following teacher directions, demonstrating self-control, making positive comments toward others, joining others in play or small groups). Whenever possible, simple-to-follow or uncomplicated strategies should be selected as they are more likely to result in intervention integrity and efficiency along with a greater percentage of adults scaffolding and supporting the behavior appropriately (Sanetti and Kratochwill 2009).

Several specific intervention strategies that have empirical support are suggested for use as tier 2 interventions, including modeling and guided practice strategies, coaching strategies, supportive and corrective feedback, peer-mediated strategies, and self-monitoring strategies (Greenwood et al. 2011; Stoiber 2004). When implementing these strategies, it is important that teachers/interventionists follow specific step-by-step procedures to assure they are conducted in a systematic manner. Stoiber (2004) suggests using the following steps when implementing modeling and guided practice: (a) determine what skill or competency will most benefit the student (e.g., select keystone behaviors leading to the student being accepted), (b) model the target behavior for the child several times, (c) provide opportunities for child to practice and rehearse the target behavior, (d) offer expanded experiences or situations in which the skill or competence can be applied and used successfully (e.g., small group, lunch room, playground), and (e) present specific feedback by stating or describing explicitly how and why the behavior was appropriate or inappropriate.

Peer-mediated strategies also have considerable empirical support (Greenwood et al. 2011; Latz et al. 2009; Stoiber 2004). Peer-mediated strategies incorporate the child's peers as models or "teachers" to support his/her development of social competencies. Peer-mediated approaches may be used, for example, to provide better or alternative ways for responding to aggression, resolving a conflict, or completing classwork assignments. Types of peer-mediated strategies include peer proximity, peer prompting, peer initiation, peer–buddy interventions, and peer tutoring.

Several structured intervention programs may be especially useful for implementation of tier 2 as a method of responding to and preventing additional problem behavior, while at the same time, teaching expected and/or alternative response behavior. One program described by Crone et al. (2010) is check in/check out (CICO). To implement it, schools provide a CICO mentor with whom selected students meet at the beginning of the day to review their behavioral expectations, identify solutions to respond to any potential barriers to appropriate behavior, practice the behavior, and review their goal for obtaining daily points. Throughout the day, the student receives feedback using a daily progress report, which is reviewed by the adult mentor at the end of the day. Upon reaching the established number of daily goal-linked points, the student may be awarded a "prize" or reward. Several researchers have documented that CICO effectively reduces problem behavior across elementary and secondary students and is endorsed by school personnel as an acceptable intervention (Todd et al. 2008). The daily progress report also may be used to communicate the students' progress toward goals with families. Caution should be exercised when incorporating this home component, however, because it may be misused by a parent and result in the child being punished. Moreover, the CICO approach should not only focus on reducing problem behaviors but also on facilitating students' development of appropriate social competencies by including goals for positive behavior to monitor on the daily progress report. In this regard, CICO can be better aligned with the intent of MTSS in helping students develop social competencies such as self-control and positive classroom behaviors which are associated with improved academic success (Durlak et al. 2011; Stoiber 2004).

Another program that might be implemented as a tier 2 intervention is *Check & Connect,* which is designed to deter drop out in at-risk elementary, middle, and high school students. Similar to CICO, a mentor meets with students who are identified as benefiting from additional support to stay in school. At daily meetings with the student, the mentor discusses the importance of staying in school and monitors the student's grades, tardiness, absenteeism, and discipline infractions. Other supports may also be provided, if needed, including a behavior plan, academic tutoring, parent counseling or consultations, and social skill groups. Researchers who designed the *Check and Connect* program have conducted several studies indicating it effectively reduces problem behaviors in students with emotional and behavioral concerns at both the elementary and secondary level (Lehr et al. 2004; Sinclair et al. 2005). An additional manualized social skills program designed to provide targeted support for students who do not respond favorably to universal class-wide programs is the *Intervention Guide of the Social Skills Improvement System* (SSIS; Gresham and Elliott 2008). The intervention guide provides instruction for teaching 20 keystone social skills within a small-group structure (1 h/week), with each skill following a modeling and guided feedback format (i.e., tell, show, do, practice, progress monitor, generalize). There also are a number of cognitive-behavioral intervention programs that may include components such as goal setting, interactive role plays or activities, behavioral contracting, and corrective feedback (e.g., Steps to Respect; Resolving Conflict Creatively Program; see http://www.whatworks.ed.gov)). For students who experience internalizing issues or school refusal behavior, programs that focus on depression or anxiety may be applied. Specifically, Stark and Kendall's (1996) *Taking Action* is designed to treat students with depression and Kendall and Hedke's (2006) *Coping Cat* is for treating anxiety.

Intensive or Tier 3 Interventions Tier 3 interventions are implemented with students who require behavior support that is highly specialized, intensive, and individualized. Within the third tier, interventions focus on teaching functionally equivalent, replacement or alternative response behaviors; placing problem behaviors on extinction; strengthening the contingencies between behavior and positive consequences; and, if necessary, applying negative consequences to eliminate severely disruptive and potentially harmful challenging behaviors (Gresham et al. 2013; Sugai and Horner 2009). Tier 3 may be delivered one on one or in small groups to the approximately 1–5% of students who do not respond sufficiently to approaches in the first two tiers. Tier 3 interventions are more strategic and focused and often of considerably longer duration than the 6–20 weeks of supplemental approaches that occur within tier 2. At tier 3, functional assessment of variables that influence student behavior is highly recommended for use in determining individualized and customized interventions (Gresham et al. 2013; Iovannone et al. 2009; McIntosh et al. 2008). Functional assessment approaches, which include examining ecological and environmental influences on the student, should not be reserved only for use at the highest tier, in that they should help facilitate a better understanding of the reasons associated with behavioral concerns at lower tiers as well. As the steps for conducting a functional assessment are available in the literature, they will not be described here in detail (see McIntosh et al. 2008; Stoiber and Gettinger 2011). It is important to note, however, that considerable evidence supports the use of functional assessment in planning a customized intervention for students who demonstrate significant social behavioral difficulties; in particular, when data are collected to define the concern and to determine the hypothesized function or intent of the inappropriate behavior (Jones and Wickstrom 2010; Gettinger and Stoiber 2006; McIntosh et al. 2008; Stoiber and Gettinger 2011). Also, information collected in the functional assessment should be drawn upon to determine appropriate goals or replacement behaviors, design the intervention, and monitor whether the intervention produced improved outcomes.

It is likely that due to the severity of the challenging behavior demonstrated in students requiring higher-tiered interventions, they will require a multicomponent support plan aimed at targeting several variables (McIntosh et al. 2009). Several researchers report the advantage of well-designed interventions matched to the hypothesized function that include clearly specified preventative, teaching, and altered response or reinforcement strategies (Iovannone et al. 2009; Stoiber and Gettinger 2011). A substantial body of research has documented the effectiveness of multiple-component, prevent–teach–respond/reinforce (PTR) interventions stemming from functional assessments that focus on improving target skills in small-group or individualized approaches (Gettinger and Stoiber 2006; Iovannone et al. 2009; Stoiber and Gettinger 2011). That is, multiple components should be included in the behavior intervention plan to address the multiple reasons linked to the behavior concern. For example, a student who exhibits severe bullying behavior will likely need a continuum of prevention and intervention strategies to teach the student how to resist engaging in verbal and physical confrontation. Relying on one or two of the interventions described in the tier 1 and 2 sections above will likely not suffice. Rather, the student may require explicit teacher-directed instruction in conflict resolution and negotiation strategies along with a mentor-facilitated behavior monitoring and behavioral contract/reward program. An important goal is to help the student develop skills in areas such as appropriate communication, self-monitoring of aggressive indicators, and de-escalation and to maintain a safe learning environment. Indeed, the adults involved in implementing tier 3 interventions also may benefit from explicit training in appropriate communication, mediation, assertiveness, and de-escalation strategies to use with the target student. Further, there are several available resources on EBI strategies and methods for selecting them that school personnel may find useful (See http://www.promisingpractices.net; Greenwood et al. 2011; Stoiber 2004; Stoiber and DeSmet 2010; Vannest et al. 2008).

Despite the level of empirical support for using functional assessments in conjunction with intervention planning and monitoring, there exists a solid body of literature indicating that educational professionals do not routinely have knowledge, skill, or experience in collecting function-based data to guide the development of positive support plans (Gresham et al. 2013; Iovannone et al. 2009). In addition, key aspects of functional assessments are not typically being implemented in schools, including collecting and using data to define the key concern, determine the function, specify a replacement or alternative response behavior, or monitor how the intervention is working (Watson et al. 2011). Thus, even when procedures such as functional assessment are implemented, many school professionals fail to consider the resulting function-based assessment data in determining replacement behaviors, developing hypotheses for the misbehavior, and designing a function-linked intervention (Gresham et al. 2013). Further, teachers are frequently not involved in the functional assessment despite often having access to essential knowledge necessary for conducting it accurately (Iovannone et al. 2009; Scott et al. 2008). Given many educators' limited role and experience with functional assessment procedures, it may not be surprising that Stormont et al. (2011) found 57 % of teachers were not sure whether functional behavioral assessments and intervention planning were provided at their school. Together, these indications suggest that functional assessment practices should receive greater attention in conjunction with MTSS.

Students with severe emotional and behavioral difficulties often require wraparound services, which include community-supported interventions by social services (e.g., child welfare) and mental health providers (Merrell and Gueldner 2010; Novins et al. 2013). For example, there exists considerable evidence for implementation of multisystemic therapy (MST) with violent and/or chronic juvenile defenders, especially when MST incorporates contingency management and intervention fidelity monitoring (Holth et al. 2011). There also is evidence that youth with ADHD symptoms benefit from psychopharmaceutical interventions, in particular, when used in conjunction with behavioral strategies (Novins et al. 2013). Other research-validated multicomponent approaches, primarily supported by single-subject studies, are often used as supplemental tier 2 or intensive tier 3 interventions (Scott et al. 2008). The collection of progress-monitoring data for determining whether and how the intervention is working should occur more frequently at tier 3, and may be necessary daily to monitor severe problem behaviors. These data should be reviewed regularly so that indicated adjustments to the intervention occur early.

Questions Regarding EBPs to Address Concerns Within MTSS

MTSS has significant flexibility that allows programs and schools to define the nature of tiered instruction along several instructional dimensions. Although this inherent flexibility promotes adoption of MTSS approaches by schools and districts, it also poses a significant challenge to evaluating MTSS applications in a controlled and systematic manner. First, educators and schools face considerable issues regarding best and/or empirically supported practices in MTSS. The majority of knowledge regarding implementation and effectiveness of MTSS stems from evaluations of interventions conducted by researchers, often with the research team providing assistance to schools, classrooms, and students. As a result, there exists limited information on the feasibility and cost of implementing EBP and MTSS in actual schools. For example, although data-based instructional decisions are considered essential for EBPs within an MTSS framework, little knowledge exists regarding several key applications by practitioners. For example, in terms of actual school-based implementation, it would be useful to develop an increased understanding of whether and what types of data sources and procedures are used, how decision-making practices stemming from these data are applied, and what guides differentiation and intervention practices in typical educational settings (Fuchs and Vaughn 2012). Perhaps most importantly, it is not known whether outcomes linked to EBP and MTSS practices are typically being evaluated and examined in schools.

Fuchs and Vaughn (2012) point out several unknowns and issues regarding best practices regarding student placement in tier 2 and tier 3. First, greater clarity is needed regarding the criteria for determining when and whether students should move from secondary to more intensive tertiary intervention. Other questions raised by Fuchs and Vaughn include: (a) should students remain in tier 2 for long periods of time (several years) if they do not meet benchmarks?, (b) how many times should students who meet benchmarks in tier 2, but then fail to keep on target at tier 1 and repeatedly need to return to the second tier, be allowed to move back and forth?, (c) should students who exhibit substantial deficiencies despite effective tier 1 instruction be placed in tier 2 when there are clear indications that they require more intensive intervention immediately?, and (d) should students remain in tier 2 or tier 3 (when tier 3 is not deemed special education) for multiple years when their markedly slow progress suggests it is unlikely they will ever catch up? Perhaps an even more essential question relates to policy decisions for determining when and whether MTSS is necessary and a "best practice" for identifying students who have learning disabilities (and require intensive, long-term intervention). In addition to the limited information regarding the validity of various MTSS approaches, little is known regarding the practical feasibility of recommended practices such as treatment integrity checks and matching type and level of intervention to student needs. Current circumstances surrounding implementation of MTSS provide strong support for increased attention to the development and dissemination of EBPs within MTSS.

Collectively, research findings demonstrate that multi-tiered interventions exert a substantial advantage for low-achieving and at-risk children. There are some indications, however, that selected subsets of students, including those students with severe deficiencies such as learning disabilities, may not benefit from diagnostic intervention trials that occur through multi-tiered forms of intervention. Rather, such students may require more intensive and sustained interventions immediately to achieve better outcomes (Fuchs and Vaughn 2012). Thus, although considerable evidence supports MTSS and EBP for achieving academic and behavioral outcomes for most learners, more research is needed to determine whether and how these positive outcomes can be applied to all learners, especially those who are at risk for academic and social–behavioral difficulties.

Summary, Future Research, and Needed Practice Directions

Across MTSS applications for literacy, behavior, or an integrated academic and behavior focus, EBPs are an integral aspect of creating a framework to promote positive outcomes on a school- or district-wide basis. In theory, when low-level problem behaviors or literacy concerns are mitigated through an evidence-based core curriculum, additional time and resources are created so educators can focus resources on more pervasive issues (Horner et al. 2010; Stewart et al. 2007). Despite the promise of MTSS for meeting the needs of all learners, questions remain regarding the quality and conclusiveness of the current evidence base.

Concerns surrounding the research support for MTSS stem from the fact that any multi-tiered system of support is actually a constellation of several EBPs implemented within separate tiers. The multiple tiers of integrated practices used to define MTSS, as well as the flexibility afforded to schools and districts in differentiating among tiers, make it difficult to evaluate MTSS in a systematic manner. On the other hand, there are indications of the importance of districts determining the best-tiered approach to use within their schools. When innovations, such as multi-tiered models, are adapted locally, there is evidence that they have a greater likelihood of being sustained (Berkel et al. 2011). Moreover, MTSS is a school-wide preventive framework, not a standard or scripted set of behavioral programs or literacy curricula. Whereas substantial research provides evidence of individual and separate components of MTSS, less is known about the effects of large-scale comprehensive applications of MTSS as an entire system (Berkel et al. 2011). Some experts worry that such a piecemeal approach is ineffective, causing them to question whether the sum of the research on individual parts of MTSS is as great as the whole (Burns 2010). At the same time, however, controlled studies of an entire MTSS model are complicated (Sugai and Horner 2010). Studies would need to account for the effect of every separate intervention within each tier, along with other critical components such as the level of professional development or a school's approach to data analysis for decision-making. Also, to the extent that community context (urban vs. suburban vs. rural) and social–economic conditions impact education outcomes in general (Dougherty Stahl et al. 2012; Durlak and Dupre 2008), a need exists to examine MTSS in terms of these more distal ecological factors as well.

Further complicating the situation are indications suggesting schools need help to implement evidence-based programs effectively and routinely. Surveys indicate, for example, that many schools fail to use evidence-based social–behavioral prevention programs (Ringwalt et al. 2009), that teachers lack knowledge of them (Stormont et al. 2011), and/or they are often implemented with poor fidelity (Sanetti and Kratochwill 2009). Thus, the scenario of limited incorporation of EBPs in schools likely is due to a variety of reasons ranging from school personnel not being aware of effective programs/strategies to schools lacking resources to implement them correctly and to monitor their effects. Evidence that EBPs are implemented at low rates in schools is a striking contrast to nearly 95 % of all schools reporting implementation of RTI at some level and 24 % reporting full implementation (Spectrum K–12 2011). These contrasting data support a wide gap between RTI and EBPs, despite federal mandates and initiatives requiring the use of EBPs in conjunction with the multi-tiered RTI and PBS.

Nonetheless, as schools continue to move toward district-wide applications of MTSS, research evaluating the benefits of comprehensive MTSS models (including an assessment of outcomes for culturally and linguistically diverse learners, as well as students with disabilities) will be a critical addition to the evidence base. It also will be important to consider ways to best operationalize MTSS so as to address the learning needs of more advanced students who not only can be easily disregarded but also should be considered in a multi-tiered model (Reis et al. 2011). Schools will benefit from the support of school psychologists and other school professionals who have research-based knowledge and practical experience in both MTSS and EBPs. Such expertise, paired with concerted efforts in completing the multiple steps in the diffusion process, will be essential to moving integrated MTSS approaches forward. In doing so, particular attention should be given to the steps linked to successful dissemination of MTSS and EBP approaches, including (a) accessing knowledge about available programs and procedures, (b) selecting strategies and programs that fit best with the school and surrounding community, (c) conducting implementation integrity checks, and (d) collecting outcome evaluation data to assess progress toward desired goals (Bernhardt and Hebert 2011; Durlak and Dupree 2008; Dougherty Stahl et al. 2012; Stoiber 2011; Stoiber and DeSmet 2010). A focus on implementation steps and ongoing assessment of the implementation climate should occur across district, school, classroom, and tiered levels in a systematic and efficient manner. These efforts should facilitate proper implementation of newly adopted strategies and programs and, moreover, ensure EBPs within multi-tiered service models are maintained and sustained as beneficial for all students for the long term.

References

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*(1), 23–133. doi:10.1007/s11121-010-0186-1.

Bernhardt, V. L., & Hebert, C. L. (2011). *Response to intervention and continuous school improvement: Using data, vision and leadership to design, implement, and evaluate a schoolwide prevention system*. Larchmont: Eye on Education.

Bond, L. A., & Hauf, A. M. C. (2004). Taking stock and putting stock in primary prevention: Characteristics of effective programs. *Journal of Primary Prevention, 24,* 199–221.

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of schoolwide Positive Behavioral Interventions and Supports

(PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31,* 1–26.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavioral Interventions, 12,* 161–179.

Burns, M. K. (2010). Response-to-intervention research: Is the sum of the parts as great as the whole? *Perspectives on Language and Literacy, 36,* 13–15.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Burns, M. K., VanDerHeyden, A. M., & Boice, C. H. (2008). Best practices in delivery of intensive academic interventions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1151–1162). Bethesda: National Association of School Psychologists.

Canter, A., Klotz, M. B., & Cowan, K. (2008). Response to intervention: The future for secondary schools. *Principal Leadership, 8*(6), 12–15.

Chitiyo, M., May, M. E., & Chitiyo, G. (2012). An assessment of the evidence base for school-wide positive behavior support. *Education and Treatment of Children, 35,* 1–24.

Collaborative for Academic, Social, and Emotional Learning. (2005). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs-Illinois edition*. Chicago: Author.

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program* (2nd ed.). New York: Guilford Press.

Crosson, A. C., & Lesaux, N. K. (2010). Revisiting assumptions about the relationship of fluent reading to comprehension: Spanish speakers' text-reading fluency in English. *Reading and Writing: An Interdisciplinary Journal, 23,* 475–494.

Coyne, M. D., Capozzoli, A., Ware, S., & Loftus, S. (2010, Spring). Beyond RTI for decoding: Supporting early vocabulary development within a multitier approach to instruction and intervention. *Perspectives on Language and Literacy*, 18–21.

Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities, 45,* 232–243.

Doll, B., Haack, K., Kosse, S., Osterloh, M., Siemers, E., & Pray, B. (2005). The dilemma of pragmatics: Why schools don't use quality team consultation practices. *Journal of Educational and Psychological Consultation, 16,* 127–155. doi:10.1207/s1532768xjep.

Dougherty Stahl, K. A., Keane, A. E., & Simic, O. (2012). Translation policy to practice: Initiating RTI in urban schools. *Urban Education, 48,* 350–379. doi:10.1177/0042085912461755.

Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41,* 327–350. doi:10.1007/s10464-008-9165-0.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students; social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82,* 405–432.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Reading Research Quarterly, 92,* 605–619.

Ervin, R. A., Schaughency, E., Goodman, S. D., McGlinchey, M. T., & Matthews, A. (2006). Merging research and practice agendas to address reading and behavior school-wide. *School Psychology Review, 35,* 198–223.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30–37.

Foorman, B. R. (2003). *Preventing and remediating reading difficulties: Bringing science to scale*. Baltimore: York Press.

Forman, S. G., Shapiro, E. S., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., Sanetti, S. M. H., & Stoiber, K. C. (2013). Implementation science and school psychology. *School Psychology Quarterly, 28,* 77–100. doi:10.1037/spq0000019.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45,* 195–203.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Gettinger, M., & Stoiber, K. C. (2006). Functional assessment, collaboration, and evidence-based treatment: An experimental analysis of a team approach for addressing challenging behavior. *Journal of School Psychology, 44,* 231–252.

Gettinger, M., & Stoiber, K. C. (2012). Curriculum-based early literacy assessment and differentiated instruction with high-risk preschoolers. *Reading Psychology, 33,* 11–36.

Graves, A. W., Duesbery, L., Pyle, N. B., Brandon, R. R., & McIntosh, A. S. (2011). Two studies of Tier II literacy development: Throwing sixth graders a lifeline. *The Elementary School Journal, 111,* 641–661.

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The response to intervention (RTI) approach in early childhood. *Focus on Exceptional Children, 43*(9), 1–22.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales*. Minneapolis: Pearson Assessments.

Gresham, F. M., Hunter, K. K., Corwin, E. P., & Fischer, A. J. (2013). Screening, assessment, treatment, and outcome evaluation of behavioral difficulties in an RTI model. *Exceptionality, 21,* 19–33.

Hammond, R. K., Campell, J. M., & Ruble, L. A. (2013). Considering identification and service provision for students with autism spectrum disorders within the context of response to intervention. *Exceptionality, 21,* 34–50.

Hartsell, K., & Hayes, S. (2012). Spotlight on MiBLSi. *Spotlight Briefs, 2*(1), 1–5.

Holth, P., Torsheim, T., Sheidow, A. J., Ogden, T., & Henggeler, S. W. (2011). Intensive quality assurance of therapist adherence to behavioral interventions for adolescent substance use problems. *Journal of Child & Adolescent Substance Abuse, 20,* 289–313.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11,* 133–145.

Horner, R. H., Sugai, G., & Anderson, C. A. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42*(8), 1–16.

Iovannone, R., Greenbaum, P. E., Wang, W., Kincaid, D., Dunlap, G., & Strain, P. (2009). Randomized controlled trial of the prevent-teach-reinforce (PTR) tertiary intervention for students with problem behaviors: Preliminary outcomes. *Journal of Emotional and Behavioral Disorders, 17,* 213–225.

Jones, K. M., & Wickstrom, K. F. (2010). Using functional assessment to select behavioral interventions. In R. Ervin, G. Peacock, E. Daly, & K. Merrell (Eds.), *Practical handbook of school psychology* (pp. 192–210). New York: Guilford Press.

Jones, R. E., Yssel, N., & Grant, C. (2012). Reading instruction in tier 1: Bridging the gaps by nesting evidence-based interventions within differentiated instruction. *Psychology in the Schools, 49,* 210–218.

Justice, L. M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools, 37,* 284–297.

Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M., & Kaufman, J. (2008). Effects of small-group reading instruction and curriculum differences for students most at risk in kindergarten two-year results for secondary-and tertiary-level interventions. *Journal of Learning Disabilities, 41*(2), 101–114.

Kendall, P. C., & Hedtke, K. A. (2006). *Cognitive-behavioral therapy for anxious children: Therapist manual* (3rd ed.). Llanfair: Workbook Publishing.

Kovelski, J. F., & Black, L. (2010). Multi-tier service delivery: Current status and future directions. In T. A. Glover & S. Vaughn (Eds.), *The promise of response to intervention: Evaluating current science and practice* (pp. 23–56). New York: Guilford Press.

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17,* 233–253.

Lane, K., Kalberg, J. R., & Menzies, H. M. (2009). *Developing schoolwide programs to prevent and manage problem behaviors: A step-by-step approach*. New York: Guilford Press.

Lane, K., Menzies, H. M., Kalberg, J. R., & Oakes, U. B. (2012). A comprehensive, integrated three-tier model to meet students academic, behavioral, and social needs. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 3, pp. 551–581). Washington, DC: APA.

Langley, A. K., Nadeem, E., Kataoka, S. H., Stein, B. D., & Jaycox, L. H. (2010). Evidence-based mental health programs in schools: Barriers and facilitators of successful implementation. *School Mental Health, 2,* 105–113. doi:10.1007/s12310-010-9038-1.

Lassen, S., Steele, M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in the Schools, 43,* 701–712.

Latz, A. O., Speirs Neumeister, K. L., Adams, C. M., & Pierce, R. L. (2009). Peer coaching to improve classroom differentiation: Perspectives from Project CLUE. *Roeper Review, 31,* 27–39.

Lehr, C. A., Sinclair, M. F., & Christenson, S. L. (2004). Addressing student engagement and truancy prevention during the elementary school years: A replication study of the check & connect model. *Journal of Education for Students Placed at Risk, 9,* 279–301.

Lindstrom, J. H. (2013). Introduction to the special issue: Response to intervention within the context of specific learning disabilities, emotional disturbance, autism spectrum disorders, and early childhood special education. *Exceptionality, 21,* 1–4.

McIntosh, K., Chard, D., Boland, J., & Horner, R. (2006). A demonstration of combined efforts in school-wide academic and behavioral systems and incidence of reading and behavior challenges in early elementary grades. *Journal of Positive Behavior Interventions, 8,* 146–154.

McIntosh, K., Brown, J. A., & Borgmeier, C. J. (2008). Validity of functional behavior assessment within an RTI framework: Evidence and future directions. *Assessment for Effective Intervention, 34,* 6–14.

McIntosh, K., Reinke, W. M., & Herman, K. E. (2009). School-wide analysis of data for social behavior problems: Assessing outcomes, selecting targets for intervention, and identifying need for support. In G. G. Peacock, R. A. Ervin, E. J. Daly, & K. W. Merrell (Eds.), *The practical handbook of school psychology* (pp. 135–156). New York: Guilford Press.

McIntosh, K., Filter, K. J., Bennett, J. L., Ryan, C., & Sugai, G. (2010). Principles of sustainable prevention:

Designing scale-up of school-wide positive behavior support to promote durable systems. *Psychology in the Schools, 47,* 5–21.

Merrell, K. W., & Gueldner, B. A. (2010). Preventive interventions for students with internalizing disorders: Effective strategies for promoting mental health in schools. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (pp. 799–824). Bethesda: National Association of School Psychologists.

Michigan Department of Education. (2012). *Integrated Behavior and Learning Support Initiative expands support around the state: MiBLSi supports student success*. Lansing: Author.

Miller, D. N., Eckert, T. L., & Mazza, J. J. (2009). Suicide prevention programs in the schools: A review and public health perspective. *School Psychology Review, 38,* 168–188.

National Early Literacy Panel. (2009). *Developing early literacy: Report of the National Early Literacy Panel* (NELP). Jessup: National Institute for Literacy. http://lincs.ed.gov/publications/pdf/NELPReport09.pdf. Accessed 13 March 2014.

Novins, D. K., Green, A. E., Legha, R. K., & Aarons, G. A. (2013). Dissemination and implementation of evidence-based practices for child and adolescent mental health: A systematic review. *Journal of the American Academy of Child & Adolescent Psychiatry, 52,* 1009–1025.

O'Connor, E. P., & Freeman, E. (2012). District-level considerations in supporting and sustaining RTI implementation. *Psychology in the Schools, 49,* 297–310. doi:10.1002/pits.21598.

Peterson, I. T., Bates, J. E., D'Onofrio, M. M., Lansford, J. E., Dodge, K. A., Pettit, G. S., & Van Hulle, C. A. (2013). Language ability predicts the development of behavior problems in children. *Journal of Abnormal Psychology, 122,* 542–557.

Promising Practices Network (PPN). http://www.promisingpractices.net. Accessed 7 Dec 2013.

Pyle, N., & Vaughn, S. (2012). Remediating reading difficulties in a response to intervention model with secondary students. *Psychology in the Schools, 49,* 273–284. doi:10.1002/pits.21593.

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal, 48,* 462–501.

Ringwalt, C., Vincus, A. A., Hanley, S., Ennett, S. T., Bowling, J. M., & Rohrbach, L. A. (2009). The prevalence of evidence-based drug use prevention curricula in US middle schools in 2005. *Prevention Science, 10*(1), 33–40.

Sailor, W., Sugai, G., Horner, R. H., & Dunlap, G. (Eds.). (2009). *Handbook of positive behavior support*. New York: Springer.

Sanetti, L. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445–459.

Scott, T. M., Anderson, C. M., & Spaulding, S. A. (2008). Strategies for developing and carrying out functional assessment and behavior intervention planning. *Preventing School Failure, 52*(3), 39–50.

Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children, 71,* 465–482.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78,* 427–515.

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79,* 839–911.

Spectrum K12 School Solutions. (2011). Response to intervention adoption survey 2011. http://www.spectrumk12.com//uploads/file/RTI%20Report%202011%20FINAL.pdf. Accessed 12 June 2012.

Stanovich, P. J., & Stanovich, K. E. (2003). *Using research and reason in education: How teachers can use scientifically-based research to make curricular instructional decision*. Washington, DC: Partnership for Reading.

Stark, K. D., & Kendall, P. C. (1996). *Treating depressed children: Therapist manual for taking Action*. Ardmore: Workbook.

Stewart, R. M., Benner, G. J., Martella, R. C., & March-and-Martella, N. E. (2007). Three-tier models or reading and behavior: A research review. *Journal of Positive Interventions, 9,* 239–252.

Stoiber, K. C. (2004). *Functional assessment and intervention system*. San Antonio: PsychCorp/Pearson Assessment.

Stoiber, K. C. (2011). Translating knowledge of social-emotional learning and evidence-based practice into responsive school innovations. *Journal of Educational and Psychological Consultation, 21,* 46–55.

Stoiber, K. C. (in press). A comprehensive framework for multitiered systems of support in school psychology. In A. Thomas & P. Harrison (Eds.), *Best practices in school psychology* (6th ed.). Washington, DC: National Association of School Psychologists.

Stoiber, K. C., & DeSmet, J. (2010). Guidelines for evidence-based practice in selecting interventions. In R. Ervin, G. Peacock, E. Daly, & K. Merrell (Eds.), *Practical handbook of school psychology* (pp. 213–234). New York: Guilford Press.

Stoiber, K. C., & Gettinger, M. (2011). Functional assessment and positive support strategies for promoting resilience: Effects on teachers and high-risk children. *Psychology in the Schools, 48,* 686–706.

Stoiber, K. C., & Vanderwood, M. (2008). Traditional service delivery vs. problem-solving consultative models: Urban school psychologists' use, needs, and priorities. *Journal of Educational and Psychological Consultation*, *18,* 264–292.

Stormont, M., Reinke, W., & Herman, K. (2011). Teachers' knowledge of evidence-based interventions and available school resources for children with emotional and behavioral problems. *Journal of Behavioral Education, 20,* 138–147.

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality, 17,* 223–237.

Sugai, G., & Horner, R. H. (2010). School-wide positive behavior support: Establishing a continuum of evidence based practices. *Journal of Evidence-based Practices for Schools, 11*(1)*, 62–83*.

Todd, A., Campbell, A., Meyer, G., & Horner, R. (2008). The effects of a targeted intervention to reduce problem behaviors. *Journal of Positive Behavior Interventions, 10,* 46–55.

Torgesen, J. (2009). The response-to-intervention instructional model: Some outcomes from a large-scale implementation in Reading First schools. *Child Development Perspectives, 3*(1), 38–40.

U.S. Department of Education, National Center for Education Statistics. (2011). *Digest of education statistics*. Washington, DC: Author. http://nces.ed.gov/programs/digest/d11/. Accessed 12 Nov 2013.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. A. (2007). Multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Vanderwood, M. L., & Nam, J. (2008). Best practices in assessing and improving English language learners' literacy performance. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 1847–1855). Bethesda: National Association of School Psychologists.

Vannest, K. J., Reynolds, C. R., & Kamphaus, R. W. (2008). *Intervention guide*. Minneapolis: Pearson.

Vaughn, S., & Chard, D. (2006, Winter). Three-tier intervention research studies: Descriptions of two related projects. *Perspectives*, 29–34.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75,* 165–183.

Vaughn, S., Wexler, J., Roberts, G., Barth, A. E., Cirino, P. T., Romain, M., Francis, D., Flectcher, J., & Denton, C. A. (2011). The effects of individualized and standardized interventions on middle school students with reading disabilities. *Exceptional Children, 77,* 391–407.

Wanzek, J., & Vaughn, S. (2011). Is a three-tiered reading intervention model associated with reduced placement in special education? *Remedial and Special Education, 32,* 167–175.

Watson, S. M., Gable, R. A., & Greenwood, C. R. (2011). Combining ecobehavioral assessment, functional assessment, and response to intervention to promote more effective classroom instruction. *Remedial and Special Education, 32*(4), 334–344.

What Works Clearinghouse. (2007). Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc. Accessed 14 March 2014.

What Works Clearinghouse. (2011). Procedures and standards handbook. Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf. Accessed 5 March 2015.

Ysseldyke, J. E., & Christenson, S. L. (2002). *FAAB: Functional assessment of academic behavior*. Longmont: Sopris West.

Preservice Teacher Education and Response to Intervention Within Multi-Tiered Systems of Support: What Can We Learn from Research and Practice?

David H. Allsopp, Jennie L. Farmer and David Hoppey

The focus of this chapter is on preservice teacher (PT) preparation and response to intervention (RTI) within multi-tiered systems of support (MTSS). For brevity, RTI is used when referencing RTI within MTSS. The chapter begins with a discussion of the need for PT preparation programs to integrate RTI. Next is a discussion of the current empirical research base specific to PT preparation and RTI. Then, suggestions for the systematic integration of RTI in PT preparation programs are offered. Suggestions are organized according to four major areas of need—(a) the need for PTs to understand RTI's/MTSS's purposes, critical components, and the roles they will play; (b) the need for PTs to be data literate; (c) the need for PTs to understand how to integrate knowledge of the content they teach, general and content-specific pedagogy, and evidence-based practices to promote positive student learning outcomes; and (d) the need for PTs to embrace the importance of involving families in the RTI process and to possess skill in doing this. The chapter concludes with final thoughts and suggestions for future research.

Some professionals might wonder why RTI should be a primary area of focus in teacher education programs. Teacher educators may question why they should consider revising their programs to integrate RTI, particularly when it will likely require them to make difficult decisions about what content should be omitted in order to make room for content related to RTI. As discussed later in this chapter, the effective integration of RTI will involve revisions not only to coursework-related content but it will also involve developing clinical field experiences with partner schools that are implementing RTI in order to provide PTs with applied experiences specific to RTI. To some, RTI is just like many other so-called educational reforms, it will be around for a while only to be replaced by the next reform. Certainly, these are all legitimate questions and concerns.

The short answer to why RTI should be integrated into preservice preparation programs is that RTI is being implemented nationwide in PK-12 education and teachers being prepared to work in these schools must have the knowledge and skills necessary to function effectively within RTI systems. Leaders in teacher preparation have emphasized the need for teachers at all levels to be competent with the skills needed within RTI (e.g., Danielson et al. 2007; Reschly 2007). The US Department of Education (USDOE) recognized the importance of improving teacher preparation in evidence-based practices including RTI by providing funding to support professional development efforts in school districts and preservice preparation programs (Danielson et al. 2007). The Individuals with Disabilities Education Act (IDEA) allows for the utilization of a student's response to scientifically based intervention as part of the specific learning disabilities eligibility determination process.

D. H. Allsopp (✉) · J. L. Farmer · D. Hoppey
Tampa, USA
e-mail: Dallsopp@usf.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_10

However, neither IDEA or the Elementary and Secondary Education Act (ESEA) specifically require schools to implement an RTI framework. However, a decision by the USDOE in 2011 to allow for flexibility in certain aspects of the law and how federal funding can be used (e.g., Title 1) provided states and school districts greater latitude in supporting RTI efforts (McInerney and Elledge 2013). The USDOE has provided millions of dollars in funding for centers to provide technical support to states and school districts in the implementation of RTI and related practices (e.g., National Center on Response to Intervention, the Center on Instruction, the National Center on Student Progress Monitoring). Additionally, many national education organizations have invested time and resources toward supporting the implementation of RTI including the National Center for Learning Disabilities (NCLD), the National Association of State Directors of Special Education (NASDSE), and the National Education Association (NEA; Prasse 2009). According to multiple indicators, it seems clear that RTI is part of the PK-12 education fabric and will be in the foreseeable future.

In contrast, teacher education is far behind PK-12 education with respect to adopting RTI as a critical component of what it does. Encouraging colleges of teacher education to move more quickly in the integration of RTI may be a daunting task. Prasse (2009) aptly recalls a colleague stating that, "changing programs and practices in universities is tantamount to attempting suicide by standing in front of a glacier" (p. 2). Change in the processes within higher education tends to happen slowly without reinforcement from outside. For many teacher educators, changing will require a major paradigm shift in how they do their work and could require external forces to make it happen in any comprehensive way (Prasse 2009).

It will be necessary that preservice preparation programs consider what is needed *presently* as they integrate RTI. However, programs will also need to consider the future and how teacher preparation curricula and practices will need to change as RTI practices evolve. Increasing the relevance of RTI for PTs will be of paramount importance to successfully integrate RTI. Teacher educators can increase the relevance of RTI through actions such as:

- Increasing clinical field experiences and developing more meaningful school–university partnerships
- Aligning certification with preparation rather than the other way around
- Basing what programs emphasize on demand in the field
- Utilizing technology to provide multiple pathways for students and multiple means of engagement in what they learn
- Preparing teachers to engage in "hybrid roles" that allow them to tap into different areas of talent and expertise instead of being confined to one role throughout their educational careers (Center for Teaching Quality (CTQ) 2013)

The Empirical Research Base

Limited research is available that investigates the impact of the integration of RTI within PT education programs. The authors conducted a search of electronic databases (i.e., Education Resources Information Center (ERIC), Education Full Text, and PsycINFO), as well as a hand search of several journals related to teacher education, special education, and learning disabilities (e.g., *Teaching and Teacher Education, Teacher Education and Special Education, Journal of Learning Disabilities*). Multiple studies addressed components of RTI (e.g., knowing and implementing evidence-based practices, using data-based decision-making). Fewer studies examined training in RTI as a whole, and those that did largely focused on in-service teachers. Four studies were located which explored the impact of the integration of RTI in PT education programs using quantitative and/or qualitative data sources. (These studies exclude program descriptions without any evaluation.) Three of these studies reviewed information from preparation programs, usually syllabi, for information on the level to which RTI was addressed within required coursework (McCombes-Tolis and Spear-Swerling 2011;

Müeller 2010; Rodriguez and Bohanon 2010). One study connected a university's clinical field experience in RTI with both candidates' and kindergarten students' learning outcomes (Hawkins et al. 2008).

Researchers investigated preparation programs generally (Müeller 2010; Rodriguez and Bohanon 2010) and targeted programs from specific disciplines (i.e., reading; McCombes-Tolis and Spear-Swerling 2011). Both the studies by McCombes-Tolis and Spear-Swerling (2011) and Rodriguez and Bohanon (2010) reviewed syllabi from multiple colleges and universities to determine if, and how, RTI was incorporated into preparation programs. Rodriguez and Bohanon also conducted interviews with faculty members. Both studies concluded RTI was not adequately covered within the preparation programs. McCombes-Tolis and Spear-Swerling found one of the 29 syllabi reviewed included progress monitoring and none mentioned RTI. Rodriguez and Bohanon reviewed 85 syllabi from colleges and universities in Illinois. Overall, RTI was not mentioned in the syllabi, and this was particularly true of early childhood education and elementary education courses where no courses included the topic. School psychology and special education courses covered RTI more with 5.4 % of special education and 12.7 % of school psychology syllabi including readings on it and 28.4 % of school psychology courses including a project or assignment about RTI.

Müeller (2010) reviewed syllabi and interviewed faculty members from six colleges and universities that were incorporating RTI in their teacher preparation courses. All the faculty members stated they had incorporated components of RTI previously, but now focused on it as a whole due to IDEA 2004. All the colleges and universities included in the study required RTI to be included in special education courses and about half included it in general education courses. Most programs infused topics into existing courses, and two schools had separate courses on RTI. Müeller found special education and school psychology programs included RTI in field experiences as well. None of the programs included in the study evaluated their candidates' learning of RTI overall, but did evaluate specific projects associated with components of RTI.

Hawkings et al. (2008) was the only study located which evaluated a comprehensive field experience on RTI. One special education teacher candidate and one school psychology candidate completed a field experience on RTI with a classroom of 23 kindergarteners. The candidates worked collaboratively with teachers and the school psychologist at their field placement. The candidates were responsible for activities at each tier of instruction, including universal screening, making data-based decisions, selecting and implementing research-based interventions, and monitoring progress of individual students in higher tiers of services. Results indicated both candidates were developing/competent or exemplary/highly competent on different areas of RTI on performance assessments related to knowledge and skills needed for RTI. Eight students who were selected for more intense services demonstrated an overall moderate intervention effect from the services they were provided. Further, the majority of students were on level for early literacy skills (87 % on phonemic segmentation fluency and 61 % on nonsense word fluency) by the end of the school year (increased from 52 % at the beginning of the year).

Linking PT Education to Current Practices Within RTI-Relevant Research: Suggestions for Teacher Educators and Policy-Makers

Despite the limited empirical base on PT education and RTI, the greater literature on teacher education, professional development, and RTI provides guidance for teacher educators and policy-makers. In this section, important considerations for teacher educators and policy-makers with respect to the preparation of PTs in RTI are discussed. First, RTI must be systematically integrated within teacher preparation programs if beginning teachers are going to be able to effectively function within RTI frameworks. This includes integrating RTI within both coursework and field experiences so that PTs have opportunities to understand and apply foundational components of RTI based on future roles. Second, an emphasis must be placed on data literacy among PTs. Beginning teachers must understand the types of data utilized to make instructional decisions and how to use data to make decisions at multiple levels (individual, class, school, district, etc.). Third, teacher preparation programs must emphasize content preparation in essential subject matter as well as how to align content knowledge and skills with understandings of the curriculum- and content-specific pedagogy. Fourth, PTs must have knowledge and skill in instructional practices that promote positive student learning outcomes across MTSS. Fifth, an emphasis on involving families in RTI is also essential. Families play a critical role in their children's education and PTs must value this fact and understand practices that effectively involve families in understanding and playing a role in RTI as it relates to their children. These five suggestions are framed around what PTs must understand and be able to do. Table 1 summarizes the five suggestions and implications for practice.

Suggestion #1 PTs must understand RTI and the roles they will play within RTI.

If schools are to implement RTI in a successful and comprehensive way, then it is critical that teacher educators systematically emphasize RTI within their preparation programs (Prasse 2009). Teacher educators need to be knowledgeable of RTI if they are going to do this effectively. A recent survey of 84 teacher educators on their knowledge of RTI found that 72% of faculty responded that they were "familiar" or "very familiar" with RTI (Schwartz et al. 2009). Differences were found in knowledge of RTI among special education, general education, and a dual specialization (special education and general education). Special education faculty demonstrated more knowledge of RTI while dual specialization faculty demonstrated greater efforts in seeking information and training on RTI compared to general education faculty. Colleges of teacher education must ensure that *all* teacher educators possess knowledge of RTI and can accurately represent it for PTs regardless of discipline or program type.

Teacher educators should embed foundational knowledge of RTI within the curriculum, including clinical experiences where teacher candidates have opportunities to function within RTI frameworks. Teacher educators should ensure that PTs understand and embrace the overall goal and purpose of RTI. Additionally, teacher educators should ensure that PTs have knowledge of the essential components of RTI and understand the potential roles they will play as beginning teachers specific to their chosen areas of focus (e.g., early childhood, elementary, middle school, high school, and special education).

Teacher Educators Should Embed RTI Within Coursework and Clinical Field Experiences The importance of imbedding RTI teaching and learning experiences in clinical settings cannot be overstated. Given that within RTI teachers need to develop pedagogical decision-making capacity and the conceptual underpinnings to do this, teacher education programs must critically evaluate the course content emphasized, assignments that focus on such skills, and clinical experiences that provide PTs opportunities to develop these important areas of professional knowledge and skill (Cochran-Smith et al. 2011). RTI content must be infused within and connect between both coursework and clinical-based field experiences where RTI frameworks are being implemented.

The preparedness of PTs to teach struggling learners is of particular importance with RTI. However, research suggests that generally PTs do not demonstrate the basic knowledge needed to do this effectively (Joshi et al. 2009; Shaw et al. 2007; Spear-Swerling and Brucker 2004). Scholars are beginning to uncover the positive effects of coursework that emphasizes instructional interventions coupled with skills in data collection and analysis. For example, Shaw et al. (2007) found that formal training in reading instruction resulted in increased literacy knowledge, improved field-based student outcomes, and influenced the attitudes of PTs toward their role as providers of literacy instruction. With respect to assessment practices, Mertler (2005) found that that PTs with specific coursework related to assessment were more readily able to determine

Table 1 Summary of five suggestions and implications for practice

Suggestion	Implications for practice
PTs must understand RTI and the roles they will play within RTI	RTI is embedded within coursework and clinical field experiences across programs (rather than a single course on RTI) The overall purpose/goal of RTI is emphasized throughout programs
	Essential RTI components are emphasized including how RTI components relate to the roles future teachers will play given their areas of certification
PTs must understand data and data-based decision-making	Assessment/data literacy is emphasized across programs
	Development of assessment/data literacy within teacher preparation programs structured so teacher candidates have ample opportunities to wrestle with the intricacies of data-based decision-making before they enter the profession
	Action research activities are embedded within programs as one vehicle for developing data-based decision-making skills
	Collaboration with school partners to develop assignments focused on inquiry around how PTs struggling students are responding to evidence-based instruction
	School administrators and university faculty need to facilitate PT access to student data so they can develop data analysis skills and practice asking important questions that can target struggling student learning needs
PTs must have knowledge of essential content	Essential content commensurate with the grade levels, subject matter, and roles that PTs will play with RTI is identified
	Essential content is systematically integrated within programs in developmentally appropriate ways
	The relationship between essential content and state/local curricula is emphasized
	Content-specific pedagogy is aligned with identified essential content and emphasized
	Concepts/skills foundational to student success (i.e., big ideas) in subject matter areas are emphasized as especially important for tier 2 and above
PTs must have knowledge and skill in instructional practices that promote positive student learning outcomes across RTI tiers	Awareness of effective instructional practices across RTI instructional tiers is emphasized for all PTs regardless of certification area
	Knowledge and skill in frontline preventative instructional practices within RTI are emphasized for all PTs (i.e., UDL, instruction responsive to students' cultural and linguistic needs, ESI, and SWPBS)
	Knowledge and skill of instructional practices related to more intensive levels of instructional support within RTI are emphasized
	Special emphasis is placed on those instructional practices that PTs will most likely need to implement based on their future roles within RTI
PTs must embrace family involvement and understand how to involve families in the RTI process	Recognize that many PTs have concerns about interacting with families
	Valuing family involvement is emphasized as an important disposition related to RTI
	Opportunities for PTs to practice research supported strategies for engaging families in the education of students and receive feedback for improvement are purposefully integrated within programs
	Opportunities for PTs to interact with families in diverse contexts are provided
	Specific opportunities for interacting with families are provided that center around key RTI components/activities

PT preservice teacher, *RTI* response to intervention, *UDL* Universal Design for Learning, *ESI* explicit systematic instruction, *SWPBS* school-wide positive behavior supports appropriate assessment tools within their field experiences compared to PTs who did not have a course on assessment. When graduates of teacher preparation programs lack foundational literacy knowledge and skill, the continued lack of proficient reading skills among struggling readers persists (Joshi et al. 2009). One way to integrate RTI into coursework and field experiences is to design field-based assignments to purposefully expose PTs to teaching struggling learners utilizing research-supported instructional practices initially learned in class. Teacher educators and school partners can provide scaffolded support and feedback to PTs in the field as they implement their instruction. These types of experiences not only help to develop PTs but also can positively influence the performance of students with whom they work (Hoffman et al. 2005).

Therefore, it is critical that teacher preparation programs purposefully structure course and field experiences so that important knowledge and skills related to RTI are learned in connected ways across courses and field experiences, including opportunities for scaffolded support (Allsopp et al. 2006; Hoppey et al. 2010; Spear-Swerling et al. 2005). Teacher educators must also incorporate the latest research within their teacher preparation programs so that beginning teachers are equipped to be effective implementers of instruction within RTI (Al Otaiba and Lake 2007; Connor et al. 2009; Hoffman et al. 2005; Mertler 2005; Moats and Foorman 2003).

Teacher Educators Should Emphasize the Purpose of RTI It is critical that PTs understand and embrace the overarching goal, or purpose, of RTI—to provide an instructional framework that accommodates the needs of all students and results in the improved achievement for all students. PTs must be taught that RTI is a purposeful process of which all teachers and related professionals are a part; that it is, first and foremost, a framework for getting the most effective instruction to all students for the purpose of preventing academic difficulties and enhancing learning outcomes whether students are below, at, or above grade level. Despite the fact that RTI has been increasingly implemented, studied, and written about, some conceptualize RTI simply as a process for identifying students with learning disabilities and primarily the responsibility of school psychologists and special education professionals (Danielson et al. 2007). Although identifying students with learning disabilities is an important area of emphasis within RTI (Bryant and Barrera 2009; Hoover and Patton 2008), this is not the primary purpose of RTI. When teacher educators affirm this misconception or simply do not emphasize RTI as important for most teachers (i.e., general education/content specific), then teacher candidates begin their teaching careers thinking that RTI is something other people do (e.g., school psychologists, special education teachers, RTI specialists). In fact, general education-/content-specific teachers may be the most important educational professionals in the RTI process, because they are the teachers from whom the vast majority of students will receive core instruction. To ensure beginning teachers understand and embrace the purpose of RTI and that all teachers are integral to the RTI process, teacher preparation programs must make communicating this a priority. Without truly understanding the intent of RTI, it is likely that beginning teachers will view RTI as a "requirement," something extra they must do but not something that is a primary aspect to their role as a teacher.

In a recent review of research on the effects of teacher education on teaching practice and retention by Cochran-Smith and colleagues (2011), the authors found that university-based teacher education programs have a strong effect on the instructional practices of beginning teachers. This effect is particularly strong when programs closely partner with schools where the context of schools is a fully integrated aspect of the teacher preparation program. However, they also found that when the school context does not embrace similar beliefs and practices, then teachers experience more difficulty implementing the practices they learned in their teacher preparation program. Interestingly, this outcome was less pronounced with respect to subject-specific instruction (e.g., reading or mathematics) compared to practices that relate to a more general stance about teaching and learning (e.g., social justice). These results are important with respect to the present discussion because RTI is as much a stance about what the goal of PK-12 education should be as RTI is a particular set of educational practices. If beginning teachers fail to truly assume ownership of this belief, then they may find it difficult to operationalize the intent of RTI when the school context does not embrace this notion or does so in a limited way.

Teacher Educators Should Emphasize Essential RTI Components and Roles Future Teachers Will Play Teacher educators must help PTs become familiar with the critical components and related practices associated with RTI and how they play out at the classroom, school, district, and state levels. Although individual teachers mostly operate at the classroom and school levels, it is important that they also understand the bigger picture, how RTI operates within their school district and their state. This is important so that teachers understand why policies are enacted that affect them and their students.

So, what RTI components and practices should teacher educators emphasize and all PTs understand and be able to do? Teacher educators should provide PTs with multiple opportunities to observe and practice critical RTI components such as data-based decision-making, implementing evidence-based core instruction as well as academic and behavioral interventions, tailoring lessons using differentiated instruction, and collaboration (Bean and Lillenstein 2012; Hoover and Patton 2008; Conderman and Johnson-Rodriguez 2009). These components and potential areas of emphasis for teacher educators are summarized in this section. A more extensive discussion of these topics is provided across Suggestions #2–5.

First, teacher educators should emphasize the development of skills in problem-solving to facilitate data-based decision-making among PTs. This includes implementing and monitoring student learning for tier 1 core curricula and identifying, implementing, documenting, and analyzing evidence-based academic and behavioral interventions in tiers 2 and above. Developing assessment skills in universal screening and progress monitoring, including ongoing curriculum-based data collection and analysis, is central to this process (Bryant and Barrera 2009; Conderman and Johnson-Rodriguez 2009; Hoover and Patton 2008). The need to emphasize data-based decision-making is discussed in more detail in the next section (Suggestion #2).

Second, teacher educators should emphasize coursework and field experiences focused on PTs implementing evidence-based core instruction and interventions. Teachers must be involved in making instructional decisions in the RTI model (Conderman and Johnson-Rodriguez 2009; Hoover and Patton 2008). In fact, many schools see RTI as a vehicle for school improvement. This is accomplished by providing high-quality core instruction that is aimed at meeting the needs of all students and, subsequently, designing and implementing evidence-based academic and behavioral interventions to meet the needs of students needing more targeted or intensive instruction. For example, in a study by Bean and Lillenstein (2012) on the changing roles of school personnel within RTI, participants indicated that "regardless of role, all personnel emphasized the importance of understanding the components of reading acquisition and being able to deliver effective instruction that was evidence based and focused on improving literacy learning for all students" (p. 494).

Third, teacher educators should better prepare PTs in instructional areas that have traditionally received less emphasis based on program area (e.g., elementary education vs. special education). For example, in a study by Conderman and Johnson-Rodriguez (2009) on the perspectives of beginning general and special education teachers about their preparation to collaborate, special educators indicated the need for a deeper understanding of core curricula while their general education counterparts suggested they wanted more intervention strategies, ongoing assessment ideas, and differentiated instruction techniques. At the core of RTI is the recognition that students have differing instructional needs and teachers need to be equipped to differentiate their instruction regardless of their instructional role.

Differentiation includes skills related to adjusting instruction and providing learning supports based on student needs (Conderman and Johnson-Rodriguez 2009; Hoover and Patton 2008). Therefore, differentiation that provides students the greatest opportunity to be successful with the core curriculum must become the centerpiece of *all* teacher education programs today. Later in this chapter, a framework for utilizing a Universal Design for Learning (UDL) approach to differentiating instruction across RTI instructional tiers is discussed as an important area of emphasis for improving the instructional capacity of PTs. A more expansive discussion is provided later in the chapter on preservice preparation and the integration of effective RTI-related instructional practices (Suggestion #4).

The development of effective collaboration skills is a fourth important RTI component that should be emphasized by teacher educators. The ability to effectively collaborate with other school personnel and families is essential within the RTI problem-solving and tiered instruction framework. This includes supporting and consulting with colleagues on building-based teams consisting of general and special education teachers as well as other school professionals (school psychologists, speech–language pathologists, reading coaches and intervention specialists). These problem-solving teams are critical in planning interventions for struggling learners and are responsible for reviewing school-wide student data, discussing successes and challenges, and suggesting ways to support teachers in their instructional efforts. Further, these collaborative teams serve as opportunities to develop teacher leadership as classroom teachers with expertise in specific content areas, interventions, or curriculum often provide ideas and information to support their peers (Bean and Lillenstein 2012). Additionally, PTs need development in effective ways to involve parents in the RTI process and communicate with them regularly about student progress. More about the importance of emphasizing ways for PTs to engage families within RTI is discussed in the penultimate section of this chapter.

Teacher candidates must understand and know how to weave these RTI components into their everyday instructional routines. This requires teachers to examine how their students are learning through the process of employing evidence-based instructional approaches, continually collecting data for the purposes of monitoring students' responses to instruction, and making instructional decisions based on these data. Collaboration using a team-based approach is essential. However, this cannot happen if teachers see their RTI role in an isolated way rather than as connected with others and the RTI process as a whole (Reeves et al. 2010). Therefore, it is imperative that teacher education programs consider the specific roles and responsibilities their graduates will need to successfully implement RTI and how they are and are not similar to traditional teachers roles in the past (Bean and Lillenstein 2012; Conderman and Johnson-Rodriguez 2009; Hoover and Patton 2008).

Suggestion #2 PTs must understand data and data-based decision-making.

The need for PTs to be data literate was briefly described earlier as one of the RTI components that should be integrated with teacher preparation programs. Due to its importance, a more expansive discussion is warranted. Within RTI, teachers are required to use student data to make effective instructional decisions that tailor instruction to meet the academic needs of all students (Barnes and Harlacher 2008; Jacobs et al. 2009). To do this, teachers document student progress across tiers by collecting and analyzing various types of student data. Teachers utilize these data to make subsequent instructional decisions about the type and intensity of intervention needed to meet each student's needs. This professional knowledge is described by some as *assessment literacy,* defined as the ability to design, select, interpret, and use assessment results appropriate for educational decisions (Quilter and Gallini 2000). The term *data literacy* is also used to characterize "the ability to examine multiple measures and multiple levels of data, to consider the research and to draw sound inferences" (Love 2004, p. 22).

To address the complexities of RTI, teacher educators can benefit from understanding in-service teachers' perceptions of how prepared they are to effectively use data to support struggling students. For instance, research suggests that while many teachers are skilled at gathering student achievement data, many teachers grapple with how to efficiently and effectively interpret data to inform their instruction (Mokhtari et al. 2007). In addition, research indicates that a teacher's ability to make instructional decisions is dependent upon his or her professional knowledge and skill of using data (Jacobs et al. 2009) as well as their ability to analyze student work, including informal assessments and student work samples (Jacobs et al. 2009; Mokhtari et al. 2007). Teacher educators must find ways to support the development of data literacy within teacher preparation programs in order to provide teacher candidates with ample opportunities to wrestle with the intricacies of data-based decision-making before they enter the profession.

Embedding action research activities within initial teacher education programs is one vehicle for developing data-based decision-making in PTs. The inclusion of action research or teacher research within field experiences is consonant with the key recommendation from the National Council for Accreditation of Teacher Education (NCATE) Blue Ribbon Report (National Council for Accreditation of Teacher Education 2010) which calls for schools, districts, and teacher preparation institutions to create authentic learning experiences and opportunities for PTs to become data literate, as well as knowledgeable about evidence-based instructional practices that develop an understanding of how to use data to guide instruction. Teacher education programs should collaboratively work with school partners to develop assignments focused on RTI using an action research framework to create space for PTs to conduct inquiry around how their struggling students are responding to evidence-based instruction (Jacobs et al. 2009). For example, at one university, general education and special education teacher candidates complete multiple inquiry or action research projects focused on the implementation of evidence-based practices tailored to meet the needs of struggling students (Hoppey et al. 2010). For special education majors, one project coupled with the final internship simulates the tier II RTI intervention process from start to finish. Through this process, teacher candidates focused instruction on a research question related to a specific group of students' needs, used data collection and analysis skills, and demonstrated their knowledge and skills (Hoppey 2013).

One outcome of this effort was that prospective teachers gained a deeper understanding of the connections between theory and practice by exploring evidence-based best practices and using data within the RTI framework in a practical step-by-step fashion (Hoppey et al. 2010; Hoppey 2013; Jacobs et al. 2009). This could not have occurred without mentor, university supervisor, and peer support that was embedded through the cycle of inquiry. This process promoted teacher candidates' ability to engage in data-based decision-making including the intricacies of moving between different instructional tiers, the multiple uses of universal screening and progress monitoring data, as well as how to effectively deliver differentiated, evidence-based instruction and interventions to meet the needs of diverse learners. Importantly, the perspectives of teacher candidates in these studies suggest that teacher education programs are positioned to help PTs recognize that problem-solving and using data is part of an effective teacher's daily routine. As a result, systematic inquiry can become embedded into emerging teachers' daily practice. Teacher preparation programs interested in cultivating an inquiry stance (Cochran-Smith and Lytle 1999; Dana and Yendol-Hoppey 2009) into teaching and student learning should consider embedding similar action research assignments within the RTI model. Action research provides a process that if routinized and habitualized helps to create a professional lens and responsibility for meeting the needs of all students. When embedded throughout a program, these action research assignments provide multiple opportunities for PTs to wrestle with the many nuances of the RTI model, while simultaneously facilitating PTs' acquisition of the knowledge, skills, and dispositions to pose questions related to student learning, data collection, and the use of evidence-based instructional strategies and interventions to improve students' learning.

However, challenges also exist. The challenge of developing effective collaborative partnerships with local schools and districts to align RTI efforts in authentic ways that align with actual practice of in-service educators is complex. For example, mentor teachers with whom the PTs work need to understand RTI as well as possess pedagogical skills in order to support PT learning. Additionally, school faculty, including mentor teachers and school administrators, need to be involved in the planning and implementation of field-based assignments that seek to develop PTs' knowledge skills and dispositions about RTI. Lastly, school administrators and university faculty need to facilitate PT access to student data in order for PTs to develop data analysis skills and encourage PTs to ask important learning questions that can target struggling student learning needs.

Suggestion #3 PTs must have knowledge of essential content.

What is "essential content?" For purposes of this discussion, essential content is defined as the foundational knowledge, understandings, and skills that students must possess in a particular content area at a particular time in the curriculum in order to be successful. For example, there is wide consensus among researchers on the essential constructs necessary for developing competency in reading (i.e., phonemic and phonological awareness, phonics, fluency, vocabulary, comprehension) and that teachers themselves must be knowledgeable of effective reading strategies and which ones are most effective for different students (e.g., National Institute of Child Health and Human Development 2000; National Reading Panel 2002). Researchers in mathematics are also beginning to develop greater consensus on the foundational constructs necessary for competency in mathematics. In 2008, the National Mathematics Advisory Panel identified critical foundations for success with school algebra including specific understandings and skills related to fluency with whole numbers, fluency with fractions, and geometry and measurement (National Mathematics Advisory Panel 2008). The Common Core State Standards (CCSS) for mathematics utilize the following common big ideas to frame K-12 mathematics content standards: operations and algebraic thinking, number and operations in base ten, and number and operations—fractions, measurement and data, and geometry (National Governors Association Center for Best Practices, Council of Chief State School Officers 2010). Additionally, through the development of the CCSS for mathematics, eight essential mathematics practices for students were identified for the purpose of developing deep levels of mathematical understanding and skill including making sense of problems and persevere in solving them, reasoning abstractly and quantitatively, constructing viable arguments and critique the reasoning of others, modeling with mathematics, use appropriate tools strategically, attending to precision, looking for and making use of structure, and looking for and expressing regularity in repeated reasoning.

Teachers must possess deep knowledge of the essential content they teach. However, effectively teaching essential content is a complex process that goes far beyond a teacher being knowledgeable of a particular subject area. Shulman (1986) and his colleagues were the first to propose that there was a unique set of subject-specific knowledge in teaching called pedagogical content knowledge. Since then, researchers have delved deeper into this area of study to try and more clearly delineate what is pedagogical content knowledge (e.g., Ball et al. 2008; Ball 1990; Carpenter et al. 2003; Grossman 1990; Hapgood et al. 2005; Ma 1999; Moats 1999; Wilson and Wineberg 1988). For example, building on Shulman's original work, Ball et al. (2008) proposed a theoretical framework for mathematics. Ball et al. (2008) describe a model that includes two overarching domains—subject matter knowledge and pedagogical content knowledge. Within the subject matter knowledge domain, three specific types of knowledge are described, common content knowledge (knowledge of mathematics needed by teachers and nonteachers alike), specialized content knowledge (knowledge of mathematics needed by teachers for teaching), and horizon content knowledge (knowledge of how mathematics topics related across the curriculum). Within pedagogical content knowledge, three additional types of knowledge are identified, knowledge of content and students (knowing about mathematics and knowing about students), knowledge of content and teaching (knowing about mathematics and knowing about teaching), and knowledge of content and curriculum (knowing mathematics and knowing the various programs and materials designed for teaching of particular subject at a given level).

Indeed, teaching subject matter is a much more sophisticated process than simply understanding the concepts and skills associated with a particular subject area, it involves knowledge and skill in other areas as well—understanding specific content for teaching in PK-12 schools, understanding how concepts and skills relate across the curriculum, understanding how students form understandings of particular concepts and skills at different developmental levels, understanding subject-specific pedagogy that aligns with the content and how students learn the content, and understanding the variety of curricular materials and programs available for content at particular levels. Frameworks such as the one described by Ball and her colleagues (2008) can help teacher education programs by providing a structure for program faculty to systematically plan how the program addresses developing PTs' content and pedagogical knowledge in ways that get at the complexity of the interrelationships among content knowledge and skills and pedagogical knowledge and skills.

With respect to preparing PTs for RTI frameworks, programs will need to be systematic in their planning around content and pedagogical knowledge and skill to ensure PTs are being prepared for the roles they will assume within RTI. Within multi-tiered instructional frameworks such as RTI, students will have different content and pedagogical needs dependent upon the level of instructional intensity they require. Across instructional tiers, teachers must be able to pinpoint students' knowledge and skill gaps in particular content areas for targeted instruction, particularly gaps related to essential content. Students who receive more intensive supports (i.e., tier 2 or tier 3 services) have greater knowledge and skill gaps than students receiving less intensive (i.e., tier 1 or tier 2 services) supports. These students will need teachers who can identify the essential concepts and skills they lack and who can pinpoint instruction to address these gaps. Dependent upon the role a teacher has within RTI, they will need to possess the combination of content, pedagogy, and student knowledge representative of the needs of students receiving that teacher's instruction. This means that teacher educators must have a clear vision for the roles they are preparing teacher candidates to assume within RTI, the essential content that must be emphasized, and how to structure their programs to ensure PTs develop the knowledge of and skills to utilize that essential content.

Suggestion #4 PTs must have knowledge and skill in instructional practices that promote positive student learning outcomes across RTI tiers.

PTs need support in developing knowledge and skill in pedagogy related to RTI. Recent research suggests that PTs lack knowledge and skill in instructional practices that are critical within RTI frameworks. Washburn et al. (2011) surveyed 91 elementary PTs to evaluate their knowledge of important constructs for teaching struggling readers. Washburn and her colleagues report that PTs demonstrated implicit knowledge and skills of some basic language constructs, such as syllable counting, but lacked explicit knowledge of others, such as phonics principles. Spear-Swerling and Chessman (2011) surveyed 142 elementary grades teachers, 98 enrolled in state university graduate programs, regarding their knowledge of practices related to RTI elementary reading. The authors found that participants lacked substantial knowledge of RTI and foundational reading assessment practices and pedagogical practices. Knowledge gaps of PTs in instructional practice related to RTI will have negative repercussions for struggling students when PTs begin teaching in schools implementing RTI.

In order for RTI to be successful, teachers must utilize instructional practices that address the learning needs of their students in response to student performance data across tiered instructional levels. How can teacher education programs support PTs to develop such knowledge and skill? As previously discussed, knowledge of essential content and understanding data and how to use data to make instructional decisions are important first steps. PTs also should possess a basic level of awareness of the different types of instructional practices that can be utilized effectively across tiers (e.g., tier 1, tier 2, tier 3) and knowledge and skill in research-supported effective practices related to the RTI roles they will play given their area of certification (e.g., early childhood, elementary, special education, secondary mathematics, secondary English, etc).

PTs Need to Have Awareness of Effective Instructional Practices Across RTI Instructional Tiers RTI is a team effort. Certainly, teachers must possess deep knowledge and skill in the practices they directly implement based on their role in the RTI process. However, teachers must also be aware of how their instructional role and the practices they implement integrate with the instructional roles and practices utilized by other teachers within MTSS. Awareness means that all teachers should have a basic level of understanding about the types of instructional support that are utilized across all tiers. Importantly, all teachers should fully understand how the types of instructional practices they provide integrate with the types of instructional practices and support provided by teachers at other tiers. For example, teachers who primarily provide tier 1 instruction should be aware of how their instruction intersects with tier 2 instruction and above. Teachers who primarily provide instruction at more intensive tiers should have awareness of the types of instruction utilized at less intensive tiers.

Problems can arise within RTI when instructional tiers are not well integrated. For example, it is critical for teachers working with the same students at different instructional tiers to communicate and understand how their combined efforts are moving the students forward toward a common trajectory. Whether instruction that occurs at tier 2 and tier 3 correlates with the core curriculum is another important area of awareness among teachers within RTI. There are times when more intensive supports should correlate with the core curriculum because supports are focused on helping students receiving tier 2 services grasp greater understandings of concepts/skills introduced in the core curriculum presently. At other times, students receiving more intensive supports will need tier 2 or tier 3 instruction on foundational concepts and skills that underlie the content being covered in the core curriculum but which may not be the specific concepts and skills being addressed presently. Therefore, teachers must be aware of and be able to complement teaching and learning practices that are utilized across tiers and between teacher colleagues.

The nature of the teaching practices utilized across tiers will differ; consequently, teachers should understand how and why teaching methods differ based on the learning needs of students. As students require more intensive supports, the level of explicitness will likely increase. This can sometimes result in differences of opinion about what "good" teaching is or is not. For example, an early grades literacy teacher may utilize more implicit types of instruction in her classroom, while a reading coach or special education teacher might employ more explicit types of teaching methods when working with small groups or individual students. Each teacher needs to understand the purposes of and the impact that different types of teaching methods have on students with differing learning strengths and needs, *instead of thinking that there is one way and one way only to teach all students*. When teacher educators are able to help instill this disposition among PTs, these teachers will more likely carry this value with them when they begin teaching. In turn, this will make it more likely that teachers will work collaboratively to adjust their instruction based on the needs of their students regardless of tier.

PTs need to possess knowledge and skill in frontline preventative instructional practices within RTI Increasingly, researchers, policy-makers, and practitioners are identifying how to most effectively integrate scientifically based practices within RTI that have potential for preventing potential academic failure for most students across tiers. Examples of these practices include UDL; instruction that is responsive to cultural, linguistic, and socioeconomic difference; explicit systematic instruction (ESI); and school-wide positive behavior supports (SWPBS). All teachers should be knowledgeable of these practices as they can be applied across instructional tiers for the purpose of preventing future learning difficulties for many students at risk for failure. Basham et al. (2010) describe an ecological framework for integrating UDL within RTI. Basham and his colleagues recommend that all teacher preparation programs integrate UDL-based instructional design in their curricula and describe how it can be comprehensively embedded within RTI. The authors identify specific areas of focus that teacher educators should consider in the preparation of leadership personnel, general education teachers, and special education teachers based on roles within RTI. For example, leadership candidates should have knowledge of leading teachers' use of effective instructional practices, knowledge of strategies for flexible scheduling, and knowledge of how to support UDL through technology. Teacher candidates should all have knowledge of effective instructional practices, knowledge of how to support student learning through technology, and knowledge of UDL. Special education teachers, in particular, should have knowledge of scientifically based individualized instructional practices including individualized use of technology (e.g., assistive technology). It is important that all teachers are able to implement instruction that incorporates UDL principles and practices within RTI because UDL provides a teaching and learning framework that can increase access to learning for all students whether they are struggling, on grade level, or advanced.

Teachers must be responsive to the diversity among students within RTI including students' diverse life experiences and prior knowledge, and how they identify themselves culturally, ethnically, linguistically, sexually, and religiously. A significant number of students in USA schools represent diverse cultures, languages, and socioeconomic status. Historically, many of these students have experienced less than satisfactory educational outcomes in US schools—e.g., students who are African American, Native American, Hispanic/Latino, English Language Learners (ELL), and students who live in poverty (Aud et al. 2013). Students who are culturally and linguistically diverse need teachers who are culturally responsive in disposition and in practice. Klingler and Edwards (2006) describe cultural considerations that educators need to make within RTI. At tier 1, teachers should utilize evidence-based practices and possess "culturally responsive attributes" (Klingler and Edwards 2006, p. 113). In order to develop the stance and skill to be culturally responsive, Klingler and Edwards describe the need for PTs to participate in experiences that prepare them to work in diverse settings (i.e., clinical experiences in schools with high percentages of culturally and linguistically diverse students that are implementing RTI). At tier 2, teachers need to be able to adapt and differentiate the types of tier 1 instructional practices students are already receiving. As students demonstrate the need for more intensive instruction, Klingler and Edwards suggest that an ELL or bilingual specialist be included as a member of the problem-solving team when making decisions about the nature of interventions and consideration of need for special education services. Socioeconomic status is a characteristic of diversity that must also be recognized and addressed within RTI (Hernandez Finch 2012). Hernandez Finch (2012) suggests that teachers should possess accurate knowledge of the factors that lead to poverty and the impact poverty has on school success for children and youth. For example, building strong relationships with families is an important skillset for teachers. Strong family–teacher relationships can help teachers understand why students may have difficulties attending school regularly (e.g., need for older siblings to stay home and care for younger siblings so adults can work, illness due to lack of health care) or why they may have difficulty with attention in class (e.g., hunger, neurological effects of untreated exposure to toxins that are more likely to be present in impoverished areas). When teachers are knowledgeable about the realities of cultural, linguistic, and economic difference and the impact they can have on the educational lives of children and youth, it is more likely that teachers will change commonly held belief systems that are negatively biased (e.g., deficit-based beliefs) and develop affirmative dispositions which in turn predispose teachers to engage in responsive pedagogical practices (Beecher 2011). In conjunction with the utilization of scientifically based instructional practices, beginning teachers will be much better prepared to effectively instruct students with diverse learning needs when they are able to utilize these principles in their teaching.

Knowledge of and skill in implementing ESI should also be a critical area of emphasis for teacher preparation programs with respect to RTI. ESI has a robust evidence base regarding its effectiveness as an instructional approach for struggling learners, whether within general education or special education contexts (e.g., Baker et al. 2002; Ehri et al. 2001; Foorman 2007; Haas 2005). Examples of ESI instructional elements that all teachers should be knowledgeable of include use of advance organizers, explicit modeling, guided practice/scaffolded instruction, independent practice, immediate positive and corrective feedback, progress monitoring, and multiple student response opportunities. ESI can be utilized regardless of instructional tier; therefore, teachers must be able to flexibly implement ESI at appropriate levels of explicitness according to students' needs, the nature of the content, and level of intensity of the intervention (Mercer et al. 1996).

PTs must also have knowledge of SWPBS and be exposed to school-based models during their field experiences. Like RTI, the emphasis of SWPBS is on prevention of behavioral difficulties. Academic success does not occur separate from social–emotional/behavioral success; therefore, within RTI, PTs must understand that practices related to behavior should not be thought of as separate from practices related to academics. When SWPBS is implemented appropriately (i.e., when all school personnel value it and follow through with its implementation), it has great potential for improving both behavioral and academic outcomes of students. Sugai and Horner (2008) describe SWPBS "as a whole-school approach emphasizing effective systematic and individualized behavioral interventions for achieving social and learning outcomes while preventing problem behaviors" (p. 69). There are several principles that guide the implementation of SWPBS that PTs should be knowledgeable of, including placing priority on the prevention of behavioral difficulties rather than only on intervention when behavior problems occur; the utilization of research-based behavior support practices; that SWPBS occurs within a comprehensive system of mental health where research-based practices are supported by building level leadership and professional development; and that data-based decision-making is utilized to guide actions (Sugai and Horner 2008). Like RTI, SWPBS incorporates a continuum of supports (tiers) based on intensity of need (e.g., whole school/whole class, small group, individualized). It is important that all teachers are able to implement the principles and practices of SWPBS because fidelity of implementation leads to decreases in rates of problem behaviors (Leedy et al. 2004). When behavior and school climate improve in schools, academic outcomes also improve (Fleming et al. 2005).

Classroom management is an essential component of SWBS. Sayeski and Brown (2011) describe a classroom management planning and implementation framework for general and special education teachers within RTI. Examples of suggested tier 1 practices include clear communication using step-by-step instructions, establishment of routines and procedures, modeling of expected behaviors, and maintaining high expectations. Examples of suggested tier 2 practices include differentiating instruction to provide more intensive supports, use of positive reinforcement systems, behavior contracts, surface management techniques like planned ignoring, proximity control, and use of humor to reduce tension. Examples of suggested tier 3 practices include functional behavior assessments, social skills instruction, self-monitoring strategies, and use of support groups (e.g., study skills, anger management).

PTs need to be knowledgeable of practices related to more intensive levels of instructional support within RTI Much research related to RTI has focused on evidence-based instructional practices related to struggling learners (i.e., students in need of intensive instructional intervention) receiving intensive instructional support (e.g., tier 2 and above). This research base can inform all teacher education programs about RTI practices that should be considered for integration within teacher preparation programs because teachers prepared in both general education and special education teacher preparation programs will have responsibility for teaching students in need of more intensive instructional intervention (i.e., tier 2). Harlacher et al. (2010) describe nine factors that should be considered for intensifying instruction within RTI: time allotted for instruction, instructional grouping, repetitions for success, amount of judicious review, interventionist facilitating the group, pacing, praise-to-corrective feedback ratio, precorrection, and error correction. Vaughn et al. (2012) describe research-supported practices for providing intensive interventions for struggling learners in reading and mathematics classrooms. The authors identify practices that relate to supporting cognitive processing including executive functioning and self-regulation, practices that relate to instructional delivery including ESI and increasing opportunities for students to respond and receive feedback, strategies for determining when and how to increase amounts of instructional time that students are receiving, and information regarding group size including information about making decisions for determining group sizes for intervention. PTs who will be responsible for tier 2 instruction and above within RTI will need to have knowledge and skill in practices such as these. For example, students with more intense needs will require more instructional time. Consequently, teachers must know how to flexibly schedule instructional time and make the best use of a finite amount of instructional time by increasing the amount of time students are actively engaged in instruction (i.e., academic learning time). Correspondingly, teachers must also be adept at planning and implementing regular opportunities for students to review and practice previously taught content (i.e., judicious review) within or outside the instructional or intervention block. PTs' abilities to implement ESI and provide struggling learners multiple response opportunities with feedback will be critical to their effectiveness in tier 2 and above contexts.

Much of the research on scientifically based instructional practice within RTI has focused on reading and mathematics interventions at tier 2 and above (e.g., Gersten et al. 2009a, b; Vaughn et al. 2012; Faggella-Luby and Deshler 2008; Scammacca et al. 2007; Newman-Gonchar et al. 2009; Fuchs et al. 2004, 2005; Bryant et al. 2008. Examples of effective literacy practices included:

- Daily intensive systematic small-group instruction on one to three foundational reading skills, per week
- Extensive one-to-one interventions (lasting at least 100 sessions)
- Explicit comprehension strategy instruction related to identifying text structures
- Discovering word meaning, self-monitoring and self-questioning, summarizing, and applying prior knowledge to make meaning of text
- Interventions that focus on direct vocabulary instruction, word study, and fluency-directed interventions (e.g., repeated reading)

Examples of effective mathematics practices included:

- Small-group instruction using a concrete–representational–abstract (C–R–A) sequence of instruction paired with computer-based practice
- Small-group instruction using a C–R–A instructional sequence that was supplemental to core mathematics instruction and that targeted foundational concepts/skills that students at risk for mathematics failure struggle with most
- Whole-class explicit instruction in using previous learned mathematics skills to problem-solve in conjunction with explication of how new problem-solving contexts fit within previously learned effective problem-solving

schemas—i.e., processes/approaches used to solve problems

- ESI that includes providing models of efficient problem-solving, verbalization of thought processes, guided practice, corrective feedback, and cumulative review
- Instruction on how to solve word problems, including teaching students the structure of problem types and how to discriminate superficial from substantive information for determining when to utilize problem-solving strategies that they have learned
- Use of visual representations of mathematical ideas (e.g., manipulatives, drawings, graphs, number lines, bar models)
- Short (i.e., about 10 min) daily sessions that attend to arithmetic fact retrieval. Exposure to evidence-based instructional practices such as these and opportunities to practice implementing them with feedback is essential within preparation programs to develop PTs' abilities to implement effective interventions at tier 2 and above

In regard to our fourth suggestion concerning PTs' knowledge of and skill with practices associated with RTI, there is a limited but burgeoning research base for effective instructional practices within RTI. This research base provides teacher educators and policy-makers a foundation for the types of practices that PTs need to develop knowledge of and skill in implementing in order to effectively serve within RTI frameworks. There is some evidence that PTs lack knowledge of important constructs and practices needed within RTI. In this section, several preventative pedagogical practices that have promise for serving as a foundation for effective instruction across tiers have been summarized (i.e., UDL, practices that are responsive to students' cultural and linguistic diversity, ESI, and SWPBS). Additionally, instructional practices that can be applied effectively for the purpose of intensifying instruction within RTI (e.g., judicious review, pacing, praise-to-corrective feedback ratio, precorrection, and error correction) have been identified and can provide PTs a frame for structuring more intensive instructional support. Examples of effective content-specific instructional practices in literacy and mathematics at tiers 2 and 3 have also been identified. Collectively, these practices are an excellent place to start as teacher educators think about their programs, the roles their PTs will play within RTI, and which instructional practices to integrate and emphasize within course and field experiences related to RTI.

Suggestion #5 PTs must embrace family involvement and understand how to involve families in the RTI process.

The need to involve families in RTI is well documented (e.g., Allsopp et al. 2010; Klingler and Edwards 2006; Sheridan and Kratochwill 2007). Moreover, family involvement can play a critical role in the improvement of students' academic and behavioral achievement in school (Harris and Goodall 2008; Sheldon and Epstein 2002). Researchers have suggested models for family–school engagement. Several have suggested models or processes that can be utilized within RTI-type frameworks. For example, Reschly (2009) identified a framework for how to develop collaborative relationships with families that corresponds with an RTI framework. Schools must (1) create opportunities and structures for family–school interactions to occur, (2) purposefully create an atmosphere that promotes open communication and shared decision-making, and (3) create an atmosphere that welcomes families and helps them to feel a part of the school community. Allsopp et al. (2010) propose a structure for schools to evaluate the extent to which they are fostering positive family engagement and to plan for improvement. Researchers have explored the beliefs and practices of PTs related to family engagement within mathematical RTI. Baum and McMurray-Schwarz (2004) studied the perspectives of PTs regarding family involvement. The authors found that teacher candidates had concerns about the nature of the relationship between teacher and parent, their abilities to meet the basic needs of children in their classrooms, and the roles that parents play in school. Graue and Brown (2003) surveyed PTs beginning their elementary or secondary teacher preparation programs. Graue and Brown report that the PTs entered their programs with ideas of family and education that were reflective of their own experiences. If PTs are not provided opportunities to expand their notions of families and how families interact with schools through opportunities to work with families in a variety of settings, the authors conclude that they will have inadequate strategies for engaging with families when they begin teaching.

Therefore, it is important that teacher educators understand the concerns of PTs so that they can work with them to develop knowledge, skills, and dispositions necessary for success in developing and sustaining positive family relationships. Additionally, teacher educators can utilize the research base related to family–school engagement, including suggestions for RTI contexts. Teacher education programs must purposefully incorporate opportunities for PTs to practice research-supported strategies for engaging families in the education of students and receive feedback for improvement. Additionally, teacher education programs must engage PTs in experiences where they have opportunities to interact with families in diverse contexts. These experiences should center around critical components related to RTI, such as communicating with families regarding students' progress-monitoring data; involving families in problem-solving and instructional decision-making; seeking information about students, their interests, their strengths, and areas of concern pertinent to learning and school success; communicating with families about the types of instruction students are receiving within RTI and reasons why the practices are being utilized; and communicating with parents about students' successes, whether great or small.

Concluding Thoughts and Suggestions for Future Research

Five suggestions for teacher educators and policymakers to consider regarding how to enhance PT preparation in RTI based on our review of the relevant research and our experiences as teacher educators who prepare PT to operate within RTI contexts have been outlined. Teacher educators will need to embrace interdisciplinary approaches to teacher education and teacher education research in order to fully implement these suggestions and push forward research efforts in this area. Just as K-12 educators must work in interdisciplinary ways for RTI to be effectively implemented, the same is true for teacher educators preparing PTs to work within RTI. Faculty in teacher preparation programs cannot operate in silos, as has too often been the case historically. Course content and the nature of field experiences will need to be better integrated to incorporate essential subject matter content, general and content-specific pedagogy, and evidence-based instructional practices that are appropriate for different instructional tiers informed by understanding student needs and student performance data. This means that teacher educators representing expertise in different areas related to RTI (e.g., elementary, secondary, special education, school and educational psychology, content experts, experts is assessment and data, etc.) will need to teach and provide field-based supervision in more collaborative ways. This includes working closely with school partners. Likely, the manner in which programs have traditionally been delivered will also need to change (e.g., co-teaching, systematically planned and blended course-field experience content delivery, school–university partnerships, etc.).

Research that directly addresses PT preparation and outcomes related to RTI is lacking. Our systematic review of the literature revealed four data-based studies on this topic. Therefore, teacher education researchers need to increase the level of research on how teacher education practices impact RTI-related outcomes. Researchers need to examine how teacher preparation around RTI impacts the knowledge and skills of PTs. Researchers also need to evaluate how such preparation impacts student outcomes during and after PTs have graduated and have begun teaching. It is hoped that the five suggestions for practice and policy provide researchers with a frame for investigating how best to emphasize the practices discussed, which areas have the greatest impact on teacher and students outcomes, and how these can be extended and improved upon based on linking RTI teacher education practice to PT and PK-12 student outcomes.

Teacher educators must be willing to work across areas of expertise inside and outside of colleges of education to better prepare PTs in RTI and to strengthen the research base for teacher education and RTI. They must be open to crossing boundaries, to learning new knowledge, and to taking risks. Transdisciplinary teaching and program development must become the norm rather than the exception in teacher education practice and research if the intent of reforms such as RTI is going to be actualized. Indeed, the quote from Prasse (2009) included in this chapter's introduction, "changing programs and practices in universities is tantamount to attempting suicide by standing in front of a glacier" (P. #2) is especially salient here. The extent to which teacher educators in higher education are willing to think and work differently will likely determine how well PTs will be prepared to effectively work within RTI/MTS in the future.

References

Al Otaiba, S., & Lake, V. E. (2007). Preparing special educators to teach reading and use curriculum-based assessments. *Reading & Writing, 20*(6), 591–617.

Allsopp, D. H., Alvarez McHatton, P., DeMarie, D., & Doone, E. (2006). Bridging the gap between theory and practice: Connecting course and field experiences. *Teacher Education Quarterly, 33*(1), 19–35.

Allsopp, D. H., Alvarez McHatton, P., Ray, S. N. E., & Farmer, J. L. (2010). *Mathematics RTI: A problem-solving approach to creating an effective model*. Horsham: LRP Publications.

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., & Zhang, J. (2013). *The condition of education 2013* (NCES 2013–037). Washington, DC: U.S. Department of Education, National Center for Education Statistics. http://nces.ed.gov/pubsearch.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*(1)*,* 51–73. doi:10.1086/499715.

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal, 90*, 449–466.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. doi:10.1177/0022487108324554.

Barnes, A. C., & Harlacher, J. E. (2008). Clearing the confusion: Response to intervention as a set of principles. *Education and Treatment of Children, 31*(3), 417–431.

Basham, J. D., Israel, M., Graden, J., Poth, R., & Winston, M. (2010). A comprehensive approach to RTI: Embedding universal design for learning and technology. *Learning Disability Quarterly, 33*(4), 243–255.

Baum, A. C., & McMurray-Schwarz, P. (2004). Preservice teachers' beliefs about family involvement: Implications for teacher education. *Early Childhood Education Journal, 32*(1), 57–61. doi:10.1023/B:ECEJ.0000039645.97144.02.

Bean, R., & Lillenstein, J. (2012) Response to intervention and the changing roles of schoolwide personnel, *The Reading Teacher, 65*(7), 491–501. doi:10.1002/TRTR.01073.

Beecher, C. C. (2011). Response to intervention: A sociocultural perspective of the problems and the possibilities. *Journal of Education, 191*(3), 1–8.

Bryant, D. P., & Barrera, M. (2009). Diane P. Bryant and Manual Barrera: Changing roles for educators within the framework of response-to-intervention. *Intervention in School and Clinic, 45*(1), 72–79.

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. M. (2008). Mathematics interventions for first- and second- grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education, 29*(1), 20–32. doi:10.1177/0741932507309712.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Portsmouth: Heinemann.

Center for Teaching Quality (CTQ). (2013). TEACHING 2030: Leveraging teacher preparation 2.0. www.teachingquality.org/teacher_prep_2.0. Accessed 9 May 2013.

Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. In A. Iran-Nejar & P. D. Pearson (Eds.), *Review of research in education* (pp. 249–305). Washington, DC: American Educational Research Association.

Cochran-Smith, M., Cannady, M., Mceachern, K. P., Piazza, P., Power, C., & Ryan, A. (2011). Teachers' education, teaching practice, and retention: A cross-genre review of recent research. *Journal of Education, 191*(2), 19–31.

Conderman, G., & Johnston-Rodriguez, S. (2009). Beginning teachers' views of their collaborative roles. *Preventing School Failure, 53*(4), 235–244.

Connor, C. M., Jakobsons, L. J., Crowe, E. C., & Meadows, J. G. (2009). Instruction, student engagement, and reading skill growth in reading first classrooms. *The Elementary School Journal, 109*(3), 221–250.

Dana, N. F., & Yendol-Hoppey, D. (2009). *The reflective educator's guide to classroom research (2nd Edition): learning to teach and teaching to learn through practitioner inquiry*. Thousand Oaks: Corwin Press.

Danielson, L., Doolittle, J., & Bradley, R. (2007). Professional development, capacity building, and research needs: Critical issues for response to intervention. *School Psychology Review, 36*(4), 632–637.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393–447. doi:10.3102/0034654307.1003393.

Faggella-Luby, M. N., & Deshler, D. (2008). Reading comprehension in adolescents with LD: What we know; what we need to learn. *Learning Disabilities Research & Practice, 23*(2), 70–78. doi:10.1111/j.1540-5826.2008.00265.x.

Fleming, C. B., Haggerty, K. P., Catalano, R. F., Harachi, T. W., Mazza, J. J., & Gruman, D. H. (2005). Do social and behavioral characteristics targeted by preventive interventions predict standardized test scores and grades? *Journal of School Health, 75*, 342–350. doi:10.1111/j.1746-1561.2005.tb06694.x.

Foorman, B. R. (2007). Primary prevention in classroom reading instruction. *Teaching Exceptional Children, 39*(5), 24–30.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students with risk for mathematics disability with and without risk for reading disability. *Journal of Learning Disabilities, 37*(4), 293–306. doi:10.1177/00222194040370040201.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J., & Hamlett, C. L. (2005). Responsiveness to intervention: Preventing and identifying mathematics disability. *Teaching Exceptional Children, 37*(4), 60–63.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009a). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools (NCEE 2009–4060)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Educational Sciences, U.S. Department of Education. http://ies.ed.gov/wwc/publications/practiceguides/. Accessed 15 Aug 2013.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009b). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009-4045)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Educational Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/. Accessed 15 Aug 2013.

Graue, E., & Brown, C. P. (2003). Preservice teachers' notions of families and schooling. *Teaching and Teacher Education, 19*(7), 719–735. doi:10.1016/j.tate.2003.06.002.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.

Haas, M. (2005). Teaching methods for secondary algebra: A meta-analysis of findings. *NASSP Bulletin, 89*(642), 24–46. doi:10.1177/019263650508964204.

Hapgood, S., Palincsar, A. S., Kucan, L., Gelpi-Lomangino, A., & Khasnabis, D. (2005). Investigating a new measure of teachers' pedagogical content knowledge for teaching informational text comprehension. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Harlacher, J. E., Nelson Walker, N. J., & Sanford, A. K. (2010). The "I" in RTI: Research-based factors for intensifying instruction. *Teaching Exceptional Children, 42*(6), 30–38.

Harris, A., & Goodall, J. (2008). Do parents know they matter? Engaging all parents in learning. *Educational Research, 50*(3), 277–289. doi:10.1080/00131880802309424.

Hawkins, R. O., Kroeger, S. D., Musti-Rao, S., Barnett, D. W., & Ward, J. E. (2008). Preservice training in response to intervention: Learning by doing an interdisciplinary field experience. *Psychology in the Schools, 45*(8), 745–762.

Hernández Finch, M. E. (2012). Special considerations with response to intervention and instruction for students with diverse backgrounds. *Psychology in the Schools*, *49*(3), 285–296. doi:10.1002/pits.21597.

Hoffman, J. V., Roller, C., Maloch, B., Sailors, M., Duffy, G., Beretvas, S. N., & the National Commission on Excellence in Elementary Teacher Preparation for Reading Instruction. (2005). Teachers' preparation to teach reading and their experiences and practices in the first three years of teaching. *The Elementary School Journal, 105*(3), 267–287.

Hoover, J. J., & Patton, J. R. (2008). The role of special educators in a multitiered instructional system. *Intervention in School and Clinic, 43*(4), 195–202.

Hoppey, D. (2013). Aligning action research and response to intervention: Developing preservice teacher knowledge, skills, and dispositions. *Networks, 15*(1). http://journals.library.wisc.edu/index.php/networks.

Hoppey, D., Morewood, A., & Bolyard, J. (2010) Response to intervention: A framework for action. In R. Pelton (Ed.) *Research in Best Practices in Action Research For Teacher Candidates*. (Chapter 8). Lanham: Rowman and Littlefield.

Jacobs, J., Gregory, A., Hoppey, D., & Yendol-Hoppey, D. (2009). Data literacy: Understanding teachers' data use in a context of accountability and response to intervention. *Action in Teacher Education, 31*(3), 41–55.

Joshi, R., Binks, E., Hougen, M., Dahlgren, M., Ocker-Dean, E., & Smith, D. (2009). Why elementary teachers might be inadequately prepared to teach reading. *Journal Of Learning Disabilities, 42*(5), 392–402.

Klingner, J. K., & Edwards, P. A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly, 41*(1), 108–117. doi:10.1598/RRQ.41.1.6.

Leedy, A., Bates, P., & Safran, S. P. (2004). Bridging the research-to-practice gap: Improving hallway behavior using positive behavior supports. *Behavior Disorders, 19*, 130–139.

Love, N. (2004). Taking data to new depths. *Journal of Staff Development, 25*(4), 22–26.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. New York: Routledge.

McCombes-Tolis, J., & Spear-Swerling, L. (2011). The preparation of preservice elementary educators in understanding and applying the terms, concepts, and practices associated with response to intervention in early reading contexts. *Journal of School Leadership, 21*(3), 360–389.

McInerney, M., & Elledge, A. (2013). *Using a response to intervention framework to improve student learning: A pocket guide for state and district leaders*. The Center on Response to Intervention at American Institutes for Research (AIR). www.air.org/files/Response_to_Intervention_Pocket_Guide_2.pdf. Accessed 23 July 2013.

Mercer, C. D., Lane, H. B., Jordan, L., Allsopp, D. H., & Eisele, M. R. (1996), Empowering teachers and students with instructional choices in inclusive settings. *Remedial and Special Education, 17*(4), 226–236. doi:10.1177/074193259601700405.

Mertler, C. A. (2005). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 33*(2), 76–92.

Moats, L. C. (1999). *Teaching reading is rocket science: What expert teachers of reading should know and be able to do*. Washington, DC: American Federation of Teachers.

Moats, L. C., & Foorman, B. R. (2003). Measuring teachers' content knowledge of language and reading. *Annals of Dyslexia, 53*, 23–45.

Mokhtari, K., Rosemary, C. A., & Edwards, P. A. (2007). Making instructional decisions based on data: What, how, and why. *The Reading Teacher, 61*(4), 354–359.

Müller, E. (2010). *Addressing response to intervention within personnel preparation programs: Six approaches. InForum*. National Directors of Special Education website: http://www.nasdse.org/publications-t577/addressing-response-to-intervention-within-personn.aspx. Accessed 8 April 2013.

National Council for the Accreditation of Teacher Education. (2010). Transforming teacher education through clinical practice: A national strategy to prepare effective teachers. http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3D&tabid=715. Accessed 8 Jan 2013.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards: Mathematics*. Washington, DC: Author.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

National Reading Panel (US), & International Reading Association. (2002). *Evidence based reading instruction: Putting the National Reading Panel report into practice*. International Reading Association.

Newman-Gonchar, R., Clarke, B., & Gersten, R. (2009). *A summary of nine key studies: Multi-tier intervention and response to interventions for students struggling in mathematics*. Postsmouth: RMC Research Corporation, Center on Instruction. http://www.centeroninstruction.org/a-summary-of-nine-key-studies-multi-tier-intervention-and-response-to-interventions-for-students-struggling-in-mathematics. Accessed 15 Aug 2013.

Prasse, D. P. (2009). The next frontier: Rolling out RTI curriculum in higher education. RTI Action Network at http://www.rtinetwork.org/rti-blog/search/entry/1/41. Accessed 23 Sept 2015.

Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*(2), 115–131.

Reeves, S., Bishop, J., & Filce, H. (2010). Response to intervention and tier systems: Questions remain as educators make challenging decisions. *Delta Kappa Gamma Bulletin, 76*(4), 30–35.

Reschly, D. J. (2007). *Overview document: Teacher quality for multitiered interventions*. National Center on Response to Intervention website: http://www.rti4success.org/resourcetype/teacher-quality-and-implementation-multitiered-intervention-system-2. Accessed 21 Sept 2015.

Reschly, A. L. (2009). Schools, families, and RTI. RTI Action Network. http://www.rtinetwork.org/essential/family/schools-familes-and-rti. Accessed 21 Sept 2015.

Rodriguez, A., & Bohanon, H. (2010). *The institutes of higher education checklist report*. Loyola University, Center for School Evaluation, Intervention, and Training website: http://www.luc.edu/cseit/pdfs/Aspire_Documents/IHE_Checklist_20100923_final.pdf.

Sayeski, K. L., & Brown, M. R. (2011). Developing a classroom management plan using a tiered approach. *TEACHING Exceptional Children, 44*(1), 8–17.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth: RMC. Center on Instruction.

Schulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Schwartz, D., Blue, E., McDonald, M., & Pace, D. (2009). Teacher educators and response to intervention: A survey of knowledge, knowledge base, and program changes to teacher preparation programs. *Journal of the American Academy of Special Education*

Professionals. *Fall 2009*, 5–19. http://www.naset.org/2903.0.html.

Shaw, D. M., Dvorak, M. J., & Bates, K. (2007). Promise and possibility—hope for teacher education: Preservice literacy instruction can have an impact. *Reading Research and Instruction, 46*(3), 223–254.

Sheldon, S. B., & Epstein, J. L. (2002). Student behavior and school discipline with family and community involvement. *Education and Urban Society, 35*(4), 4–26.

Sheridan, S. M., & Kratochwill, T. R. (2007). *Conjoint behavioral consultation: Promoting family-school connections and interventions* (2nd ed.). New York: Springer.

Spear-Swerling, L., & Brucker, P. O. (2004). Preparing novice teachers to develop basic reading and spelling skills in children. *Annals of Dyslexia, 54*(2), 332–364.

Spear-Swerling, L., Brucker, P. O., & Alfano, M. P. (2005). Teachers' literacy-related knowledge and self-perceptions in relation to preparation and experience. *Annals of Dyslexia, 55*(2), 266–296.

Spear-Swerling, L., & Chessman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading & Writing, 25*(7), 1691–1723. doi:10.1007/s11145-011-9338–3.

Sugai, G., & Horner, R. H. (2008). What we know and need to know about preventing problem behaviors in schools. *Exceptionality, 16*, 67–77. doi:10.1080/0932830801981138.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics: A practice guide*. Portsmouth: RMC Research Corporation, Center on Instruction.

Washburn, E. K., Joshi, M., & Cantrell, E. B. (2011). Are preservice teachers prepared to teach struggling readers? *Annals of Dyslexia, 61*, 21–43. doi:10.1007/s11881-010-0040-y.

Wilson, S. M., & Wineberg, S. (1988). Peering at history through different disciplinary lenses: The role of disciplinary perspectives in teaching history. *Teachers College Record, 89*, 525–539.

Common Core State Standards and Response to Intervention: The Importance of Assessment, Intervention, and Progress Monitoring

Shane R. Jimerson, Rachel Stein, Aaron Haddock and Reza Shahroozi

> As challenging as it must have been to write and finesse the adoption of the Common Core State Standards, that accomplishment is nothing compared to the work of teaching in ways that bring all students to these ambitious expectations. The goal is clear. The pathway is not. (Calkins et al. 2012, p. 4)

The K–12 educational context has continued to experience numerous shifts in policy and perspective over the decades. Increasingly, the focus has been on how to hold all students to high standards and how to accurately and fairly measure student progress. In order to help students and schools meet these new requirements, response to intervention (RTI)/ multi-tier system of supports (MTSS) practices have gained in popularity (Jimerson et al. 2007). Throughout this chapter, the acronym RTI is used to refer to RTI/MTSS practices. RTI aims to facilitate student achievement through early assessment and identification of student strengths and needs, and offering a continuum of services to support students' academic growth. With the recent debut of the common core state standards (CCSS; National Governors Association Center for Best Practices, Council of Chief State School Officers 2010, hereafter cited as NGACBP/CCSSO 2010), it is important to consider the potential opportunities and implications of RTI/ MTSS and CCSS in facilitating the academic success of all students.

S. R. Jimerson (✉) · R. Stein · A. Haddock · R. Shahroozi
University of California, Santa Barbara, CA, USA
e-mail: Jimerson@education.ucsb.edu

This chapter delineates the central tenets of the CCSS and explores intersections with RTI practices. The chapter begins with a brief overview of standards-based accountability and reform efforts in the USA during the past few decades. With this context in mind, the CCSS for English language arts and mathematics are described and critically assessed. Subsequently, the chapter explores key considerations for teacher education, professional development, and assessment related to the new CCSS standards. Building on this foundation, the chapter introduces the RTI/MTSS framework, explicating important intersections between RTI and CCSS and elucidating implications for both general and special education. The chapter concludes with practical recommendations for integrating RTI and CCSS, a discussion of common concerns and critiques, and an overview of areas for future research.

Standards-Based Accountability and Reform

The CCSS represents the next step in the standards-based reform movement in American education. The standards-based reform movement began in the 1980s, and gained momentum following the publication of *A Nation at Risk: The Imperative for Educational Reform* in 1983 (Gardner et al. 1983). *A Nation at Risk* stressed that lax educational standards were undermining the US dominance in "commerce, industry, science, and technological innovation" (National

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_11

Commission on Excellence in Education 1983). In response to these concerns, states and school districts were compelled to articulate academic standards to guide instruction and began to develop policies that based promotion and placement decisions on standardized test scores. Applebee concludes that "The Common Core State Standards (CCSS) […] are the culmination of at least 25 years of emphasis on systemic school reform, using high-stakes assessments as a lever to improve the achievement of American school children" (Applebee 2013, p. 25).

The first attempts at standards-based school reform were voluntary initiatives led by groups such as the National Council of Teachers of Mathematics in 1989 (NCTM 1989), although ultimately unsuccessful. These early efforts were followed in the early 1990s by brief, though failed attempts, by the federal government to develop national academic standards in English language arts and history. With persistent concerns about academic achievement levels in the USA, both the Clinton and Bush administrations (1993–2009) consistently pushed for more rigorous standards-based accountability (Picklo and Christenson 2005). The bipartisan legislative efforts to use standards and high-stakes assessment to reform the education system gained traction at the state level with the passage of the No Child Left Behind (NCLB) Act in 2001 (20 U.S.C. § 6301).

NCLB emphasized standards-based educational reform and stricter accountability through the assessment of measurable goals aimed at improving educational effectiveness and closing the achievement gap between diverse groups of students. The NCLB Act held students and school systems accountable to meeting content-based standards by connecting schools' performance on high-stakes tests to federal funding. In addition, not meeting the standards triggered a series of actions intended to remediate low-performing schools.

Within the context of NCLB states created their own academic standards, essentially resulting in 50 different visions of what students should know and 50 different assessment systems (Applebee 2013). Critics argued that these standards were too broad in some states, not rigorous enough in others, inadequately informed by empirical research, and ultimately unsuccessful. For example, despite years of NCLB reform efforts, in a 2009 comparison with students internationally, US students scored 14th in reading, 17th in science, and 25th in mathematics performance on the program for international student assessment (PISA), which assesses the academic achievement of 15-year-olds in 65 developed nations (King and Jones 2012). Furthermore, the PISA assessment revealed that the highest achieving students in the USA performed below other countries' top performers and that there are more low-performing students in the USA than in other developed countries (King and Jones 2012). Thus, in part, the CCSS were created to respond to this problem by raising academic achievement standards for all American students to international benchmarks. With input from the states and teams of teachers, the CCSS were designed to provide a core set of academic standards to guide academic instruction in all states.

The CCSS, released in June of 2010, represent the most contemporary academic standards held in common by the vast majority of states in the USA. The National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO) developed the standards with guidance from content experts, teachers, and researchers as well as input from parents, school administrators, civil rights groups, and other interested citizens via two public comment periods (NGACBP/CCSSO 2010). The development process began by defining the skills and knowledge necessary for success in college and the workplace and articulating them in the *college and career readiness standards*. The *college and career readiness standards* then constituted the student outcomes sought by the K–12 educational process, which in turn guided the construction of the grade-level standards (NGACBP/CCSSO 2010). As will be discussed at greater length, the evidence base linking the grade-level standards to the *college and career readiness* standards is unclear. Both sets of standards were developed by separate committees of educators and subject-matter experts, utilizing criteria created by representatives from Achieve, ACT, the College Board, the National Association of State Boards of Education, and the State Higher Education Executive Officers (Alliance for Excellent Education 2013). In relation to sets of standards found in many states, the authors of the CCSS sought to create a streamlined set of standards comprising fewer, clearer, more rigorous standards delineated in a logical progression from grade to grade.

Unlike NCLB, the CCSS initiative is a state-led effort that is not part of NCLB. The nation's governors and education commissioners, through their representative organizations the NGA and the CCSSO, led the development of the CCSS and continue to lead the initiative. It is not mandatory for states to adopt the CCSS, although the federal government has tied significant resources to adoption of the CCSS. For example, states that adopted the CCSS were evaluated significantly more favorably in the race to the top (RTT) funding competition (Alliance for Excellent Education 2013). As of 2014, the CCSS had been adopted in 45 states, the District of Columbia, four of the US territories, and the Department of Defense Education Activity (NGACBP/CCSSO 2010), whereas most states adopted the CCSS for all grade levels at one time, some states have opted to implement the new standards in a gradual process beginning in the earlier grades. Although most states will now share a common set of academic standards, they will retain decision-making power over the curricula adopted to prepare students for the CCSS tests. While the CCSS identifies the standards at each grade level, the district and school select the curricula teachers will employ and define the sequence of instruction and instructional methods employed. Thus, the standards define what students should learn and the outcomes sought, but not how that content should be delivered (commonly known pedagogy and curriculum; Porter et al. 2011). New assessments, aligned with the CCSS and the pedagogy that informs it, are slated for administration for the first time during the 2014–2015 school year.

According to the CCSS initiative, the new standards are evidence based, more rigorous than most states' current standards, focused on a smaller number of critical subjects, and designed to impart the skills and knowledge most needed for college and career readiness ("Mission Statement," n.d). Curriculum, pedagogy, and assessment changes accompany the new standards.

While the authors of the CCSS state that the standards are "evidence-based," this is misleading. Describing the standards as "evidence-based" suggests that the standards have been operationalized and subjected to rigorous empirical investigation, which is not the case. In reality, the CCSS are grounded in a mélange of research, predecessor policy documents, previous state standards, input from the team of developers, and feedback from stakeholder organizations and the public. As such, it might be more accurate to describe the standards as fairly "research based." As a result, the standards reflect areas of both consensus and dispute.

The CCSS documents include a "Sample of Works Consulted" for the mathematics standards and a "Bibliography" for the ELA standards (NGACBP/CCSSO 2010), both of which make mention of research and predecessor policy documents; however, it is unfortunate that the authors neglected to adequately clarify the research on which the standards are based. In an article on the CCSS entitled "Common Core Standards: The Emperor Has No Clothes, or Evidence" (2011), Tienken argues that the authors of the standards failed to present adequate evidence for both the particular set of standards offered and that a system of national standards is an improvement over a system of state standards. Tienken (2011) points out that although on the official web site of the CCSS it is claimed that the "*Standards* have made careful use of a large and growing body of evidence," only two documents are presented to verify this claim, a two-paragraph FAQ piece entitled "Myths v. Facts about the Common Core Standards: Myths About Process" and a report called *Benchmarking for Success* (NGA/CCSSO/Achieve 2008). Notably, the *Benchmarking for Success* report was authored by the same groups that authored the standards and is largely based on a single study by Hanushek and Woessmann (2008) that, Tienken (2011) writes, "has been criticized exhaustively and shown to be fatally flawed by independent researchers" (Tienken 2011, p. 59).

The lack of research cited in the CCSS documents seems particularly surprising given the scores of content experts that contributed to their creation and in light of the oft-heard claim that the standards are evidence based. Ultimately, this discrepancy may hinge on researchers' and policymakers' differing definitions of what constitutes an evidence base. As Cuban (2010) has clarified:

> Policymakers…use evidence…to argue that their hunch ("common core" standards) about solving a pressing problem…is both logical and compelling. Evidence that contradicts the policy is tossed aside. In short, facts or the totality of available evidence do not determine policy. Evidence is made, not found. Evidence is used to fit the hunch, arguments, and logic of a policy aimed at solving a problem.

The concern is that the current body of standards consists of research-based standards and learning progressions intermingled with others that are not research-based; yet, the CCSS are deceptively touted as evidence based. In the introduction to the ELA standards, the authors admirably explain that the CCSS are "intended to be a living work: as new and better evidence emerges, the Standards will be revised accordingly" (NGACBP/CCSSO 2010). Unfortunately, the authors' failure to clearly explicate the particular research that informs the standards and pedagogical shifts stemming from the new standards may seriously hamper such a well-intentioned project.

Common Core State Standards

English Language Arts

The ELA standards set objectives in reading, writing, speaking and listening, and language for students in grades K–12. The standards also delineate requirements for literacy in history/social studies, science, and technical subjects. The ELA "anchor standards" are based on the *college and career readiness standards,* which were developed first. The authors offer a portrait the student prepared for college and career, which defines the outcome engendered by mastery of the standards: "the literate individual" (NGACBP/CCSSO 2010). Students prepared for college and career demonstrate independence; build strong content knowledge; respond to the varying demands of audience, task, purpose, and discipline; comprehend as well as critique; value evidence; use technology and digital media strategically and capably; and come to understand other perspectives and cultures (NGACBP/CCSSO 2010).

The ELA standards are divided into three sections: a comprehensive K–5 section and two content area-specific sections for grades 6–12, one for ELA and one for history/social studies, science, and technical subjects. The reading standards are divided into standards for literature, informational text, and foundational skills. There are ten standards for reading, ten for writing, six for speaking and listening, and six for language. The reading standards emphasize both the students' developing reading skill and comprehension as well as the increasing complexity of the texts they read. The writing standards stress the development of core writing skills (e.g., planning, revising, editing, and publishing); an ability to write different types of text (arguments, informative/explanatory texts, and narratives); and writing in response to reading, including using evidence from literary and informational texts. The speaking and listening standards focus on the development of flexible and broadly useful oral communication and interpersonal skills, such as collaborating with others, express and listen carefully to others' ideas, and the integration of information from a variety of sources and mediums. Finally, the language standards delineate the conventions of standard written and spoken English, including the effective use of language and vocabulary development.

The literacy (reading, writing, speaking, and listening) standards are aimed at preparing students for college and career. Along with changes in the actual educational objectives, the CCSS further embody shifts in thinking about how to teach English language arts and reading (commonly called ELA). Importantly, the CCSS distribute the responsibility of facilitating students' literacy skills and make subject-specific reading and writing instruction a key piece of each of the core subjects (Shanahan and Shanahan 2012). The standards in grades K–5 focus on ensuring students learn to read at increasing levels of complexity as they matriculate across grades. Haager and Vaughn (2013) conclude that the K–5 CCSS are likely not much different than most states' previous *reading/language arts* standards; however, they do note that one key difference lies in the relative difficulty of grade-level texts students will be expected to read, with the CCSS' expectations being substantially more challenging. In general, the CCSS ELA standards emphasize "close readings" (i.e., rereading, interpreting perspective, and providing additional text sources as evidence to support ideas) of dense and complex informational text and content rich nonfiction, such as historical and scientific texts. The standards in the early grades focus on what the CCSS authors call *foundational skills,* or the critical early reading skills (e.g., concepts of print, the alphabetic principle), building toward the development of reading fluency, reading comprehension, and a mature vocabulary. Although students will continue to read literature and poetry in English classes, the CCSS do call for increased focus on literary nonfiction, especially in grades 6–12, yet the CCSS expect at least half of student reading to be expository or informational text by 4th grade. By 12th grade, students are expected to devote 70 % of their reading across the grade (i.e., in all classes, not just English language arts) to informational text (NGACBP/CCSSO 2010). Likewise, history/social studies and science classes will put more focus on teaching students how to extract disciplinary knowledge by reading relevant texts. Furthermore, in order to build students' comprehension of increasingly complex texts, the CCSS encourage regular practice with academic vocabulary, which Foorman defines as "the language of disciplines, texts, and of discourse" (Foorman 2012). Naturally, these more rigorous expectations will pose difficulties for students already struggling to meet grade-level expectations. For more information on how general and special education teachers can best support students with learning disabilities, meet the CCSS for ELA see Haager and Vaughn 2013.

In writing, the CCSS provide a clear learning progression of skills and applications within a more rigorous and encompassing framework (Graham and Harris 2013). The CCSS identify four key applications of writing skills: (1) learning to write for a variety of purposes, including argument, informative/explanatory texts, and narratives (*text types and purposes*); (2) producing and publishing clear and coherent text appropriate to task, purpose, and audience by planning, revising, editing, and collaborating with others (*production and distribution of writing*); (3) utilizing writing to research, integrate information, analyze, reflect, and demonstrate understanding of a topic or materials read (*research to build and present knowledge*); and (4) writing routinely over both extended and shorter time frames to produce numerous pieces across a diversity of tasks, purposes, and audiences (*range of writing*). These writing applications are supported by the Language standards, which specify skills that undergird fluent writing skills, such as sentence construction, handwriting (or typing), spelling, grammar, conventions, and word choice. The CCSS in ELA recognize the development of writing skills as on par with reading, and assign writing an equal number of individual standards. In order to prepare students for college and career, the CCSS put less emphasis on personal narrative, opinion, and memoir in favor of having students write pieces that employ factual evidence from texts they have read to inform and persuade. The authors state that an important goal of the writing standards is to foster students' appreciation of writing as a means of communicating with others that requires clarity in order to accomplish one's purpose. As Coleman (2011), a chief architect of the English language arts standards, explained in a presentation on the CCSS at the Teach for America 20th Anniversary Summit, the CCSS will ask students to "read like detectives and write like investigative reporters."

Based on what teachers commonly report on how writing is taught, the CCSS' vision for writing instruction in grades K through 12 represents a significant shift. Graham surveyed elementary and secondary school teachers about their writing instruction practices. Subsequent analyses found that elementary students do not spend much time writing each day at school (an average of 20–25 min) and do not receive adequate instruction in writing after third grade (Cutler and Graham 2008; Gilbert and Graham 2010). Secondary school teachers reported that writing was taught infrequently, primarily consisted of short responses and brief summaries, and was not taught across subjects (Applebee and Langer 2011; Kiuhara et al. 2009). In addition, it appears that most teachers believe that they do not possess the skills required to teach writing well (Gilbert and Graham 2010; Kiuhara et al. 2009).

Given the current state of writing instruction in the USA, it is clear that achieving the rigorous and lofty standards set forth in the CCSS will require significant professional development for teachers and significantly more time and attention devoted to writing instruction and practice for students. For an excellent examination of how to help students with learning disabilities achieve the *common core's* more challenging writing standards, see Graham and Harris (2013). In the article, the authors discuss four critical recommendations: (1) increase general and special education teachers' knowledge about writing development; (2) develop a writing environment where students with learning disabilities can thrive; (3) implement evidenced-based writing practices for all students in the general classroom; (4) and implement evidence-based writing practices that have been shown to work with students with learning disabilities. Furthermore, the article provides a table identifying evidence-based writing practices found to be effective in the general classroom. Finally, the authors direct readers in need of more thorough descriptions of particular interventions to the book *Best Practices in Writing Instruction* (2nd ed.) (Graham et al. 2013).

According to the Myths v. Facts document, the English language arts standards are grounded in the National Assessment for Educational Progress (NAEP) frameworks, which NAEP claims are evidence based, although no further substantiation is provided. The NAEP frameworks describe the assessment objectives and design for national content-area tests and are used to make state-to-state student performance comparisons (Jago 2009). Unfortunately, these documents do not present the wealth of references to high-quality empirical research one would expect to support the claim that the standards are based on "a large and growing body of evidence" (NGACBP/CCSSO 2010, p. 3). As Applebee (2013) has explained, "Given the speed with which the standards were written and the many layers of input from stakeholders, the document that has resulted contains residues of all of our professional disagreements about the teaching of the English language arts. Whatever you consider most important in the teaching of reading and writing, you can find it somewhere in the standards and its accompanying documentation. And so can those who disagree" (p. 26). For instance, while Applebee (2013) recognizes the many strengths of the ELA standards (e.g., their accurate representation of the consensus on reading development; the distribution literacy development across the disciplines) and concedes that the specific standards "reflect a broadly shared vision" of the English language arts, he nevertheless expresses concern regarding the standards' separate emphasis of foundational skills, the grade-by-grade standards, the lack of a developmental model for writing, and issues of implementation. Other ELA experts have expressed similar and further reservations about the ELA standards (see, for example, Bomer and Maloch 2011; Newkirk 2013; Powell et al. 2013). For example, Newkirk (2013) finds the standards' approach to developing readers capable of handling complex texts misguided, the grade-level target student texts developmentally inappropriate, the view of reading sterile, and the approach to teaching writing fundamentally flawed. Graham and Harris (2013) applaud the writing benchmarks as providing a useful and needed guide to writing instruction across the grades, but caution that the benchmarks are all too often "simply educated guesses as to what students' should be able to achieve at particular grades" that risk impeding teachers' ability to differentiate instruction for students' unique needs and developmental level (p. 31).

Mathematics

In 2001, the authors of the National Research Council's report *Adding It Up* characterized the US elementary and middle school mathematics curriculum as "shallow, undemanding, and diffuse in content coverage" in comparison to the curricula of other high performing countries (National Research Council 2001). In response, the authors of the CCSS in mathematics aimed to create a more focused and coherent body of academic standards to facilitate students' mastery of the most important skills and knowledge ("Common Core State Standards for Mathematics," n.d; Schmidt et al. 2002). In contrast to the ELA standards, the mathematics standards appear to reflect a greater consensus among experts in mathematics and mathematics education. The CCSS have benefited from efforts on the part of researchers and policy groups over the past two decades to articulate a set of key skills for proficiency in mathematics. As VanDerHeyden and Alsopp (in press) clarify in a forthcoming article, these aforementioned efforts led to the creation of a series of policy documents that, when taken together, offered a streamlined and research-based set of essential skills foundational for mathematics competency (VanDerHeyden, forthcoming). These policy documents included the National Council of Teacher's of Mathematics' (NCTM) publications: Curriculum and Evaluation Standards for School Mathematics (NCTM 1989), Principles and Standards for School Mathematics (NCTM 2000), and the Curriculum Focal Points documents (NCTM 2006). According to Alsopp and VanDerHeyden (forthcoming), these documents informed the National Mathematics Advisory Panel (2008) and, in turn, the CCSS in mathematics (2010, available at http://www.corestandards.org/Math). The CCSS in mathematics are structured around eleven content domains:

1. Counting and cardinality
2. Operations and algebraic thinking
3. Number and operations in base ten
4. Number and operations—fractions
5. Measurement and data
6. Geometry
7. Ratios and proportional relationships
8. The number system
9. Expressions and equations
10. Functions
11. Statistics and probability

After assessing the CCSS in relation to the aforementioned research-based policy documents, VanDerHeyden and Alsopp (in press) concluded that the CCSS in mathematics represent detailed and sensible outcomes to guide instruction and assessment. Similarly, following a review of the new standards in relation to research on mathematics education, Cobb and Jackson (2011) concluded that the CCSS for mathematics represent an improvement over most state mathematics standards. They specifically found the new mathematics standards more focused (i.e., fewer core mathematical ideas are taught at each grade) and more coherent (i.e., based on logical learning progressions regarding how students develop and organize their knowledge of particular mathematical domains and achieve procedural fluency and skills). In contrast, an analysis by Porter et al. (2011) concluded that the CCSS "are somewhat more focused in mathematics but not in ELAR" (p. 114). For further discussion of the CCSS in relation to best practices in mathematics education, please see VanDerHeyden and Alsopp (in press).

When developing the mathematics standards, the authors state that they paid close attention to "research-based learning progressions detailing what is known today about how students' mathematical knowledge, skill, and understanding develop over time" (Common Core State Standards Initiative 2010, n.p.). Confrey et al. (2010) analyzed the extent to which the individual standards are organized into coherent learning progressions leading toward the central mathematical ideas and concluded that the CCSS in mathematics do in fact offer a sophisticated and logical progression. In general, the shift to the CCSS standards will mean higher levels of cognitive demand and emphasis on both developing students' conceptual understanding as well as their fluency. Fluency, defined as accuracy plus speed, consists of both procedural and computational components (Binder 1996; as cited in VanDerHeyden, forthcoming). The important reciprocal interrelation between fluency and conceptual understanding is widely acknowledged among mathematics experts and explicitly endorsed in the NRC report of 2001. As such, the CCSS in mathematics expect students to achieve mastery of particular foundational concepts and procedural skills at each grade level (NGACBP/CCSSO 2010). Porter et al. (2011) analyzed the level of cognitive demand required by the new mathematics standards in relation to the average of the state mathematics standards and concluded that the CCSS in mathematics represent a modest increase in cognitive demand as well as increased emphasis on number sense and operations in grades 3–6 (55%) when compared to the state aggregate (29%).

Consistent with these changes, the CCSS also reflect contemporary evidence on best practices associated with the teaching of mathematics (Cobb and Jackson 2011). Modeled on mathematics pedagogy in high-performing countries as well as ways of thinking that research indicates proficient users of mathematics possess (Ma 1999), the CCSS seek to facilitate students' depth of understanding and procedural fluency with the *standards for mathematical practice:*

> The Standards for Mathematical Practice rest on important "processes and proficiencies" with longstanding importance in mathematics education. The first of these are the NCTM process standards of problem solving, reasoning and proof, communication, representation, and connections. The second are the strands of mathematical proficiency specified in the National Research Council's report *Adding It Up*: adaptive reasoning, strategic competence, conceptual understanding (comprehension of mathematical concepts, operations and relations), procedural fluency (skill in carrying out procedures flexibly, accurately, efficiently and appropriately), and productive disposition (habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy). Common Core State Standards for Mathematics, n.d., p. 6)

Taken together, the CCSS offer mathematics educators the following eight mathematical practices:

1. Make sense of problems and persevere in solving them
2. Reason abstractly and quantitatively
3. Construct viable arguments and critique the reasoning of others
4. Model with mathematics
5. Use appropriate tools strategically
6. Attend to precision
7. Look for and make use of structure
8. Look for and express regularity in repeated reasoning

Students are expected to develop proficiency in the *standards for mathematical practice,* and the practice standards will be assessed along with the content standards.

Experts on mathematics education agree on the reciprocal nature of conceptual understanding and fluent skill performance in mathematics learning (NMP 2008; Wu 1999). As articulated in the NRC (2001), "Understanding makes learning skills easier, less susceptible to common errors, and less prone to forgetting. By the same token, a certain level of skills is required to learn many mathematical concepts with understanding, and using procedures can help strengthen and develop understanding" (p. 122). As a result, the CCSS in mathematics place equal emphasis on conceptual understanding and procedural fluency, which represents a significant departure from current practice in most districts and schools. Research consistently indicates that early childhood teachers are often less skilled in the teaching of mathematics relative to reading, believe literacy is more important than mathematics, and spend less classroom time on mathematics activities (Early et al. 2005; Layzer 1993; Hausken and Rathbun 2004; Musun-Miller and Blevins-Knabe 1998). For instance, Ma's (1999) research reveals that although US teachers typically have had 4–7 more years of formal schooling than Chinese teachers, US teachers have a less well-developed understanding of elementary mathematics. High teacher performance is dependent upon sophisticated subject-matter knowledge, clear understanding of the progression of skill development, and competency in lesson design and delivery, including facility with multiple approaches to demonstration and explanation. As VanDerHeyden and Alsopp (in press) summarize, "The most powerful lesson of the research data is that content knowledge (knowing what to teach) and the science of effective instruction (knowing how to teach) are both essential prerequisites to successful learning outcomes for students in mathematics" (p. 7; also see for example Hattie 2013 and Slavin and Lake 2008). Ma (1999) shows how the US teachers' weaker understanding of both the mathematics they teach and mathematics pedagogy leads to diminished outcomes for their students.

While helping teachers develop their understanding of the new mathematics standards is fairly straightforward, facilitating their skills in a different approach to instruction will likely be more challenging. As a result, teachers of mathematics often require more specific and strong curricula, when compared to reading, but research reveals mathematics curricula are weak compared to reading and poorly aligned with the tenets of instructional design (VanDerHeyden and Alsopp, in press). Moreover, mastering the new mathematics standards may pose difficulties for some teachers since the vast majority have never experienced this approach to instruction, either as students or teachers in training, and may not possess the level of mathematical understanding required to help their students be successful (e.g., Ma 1999). At present, students in the USA are typically taught a method (or algorithm) to solve a particular kind of problem and then asked to repeatedly apply that method to problems of increasing difficulty. By contrast, the CCSS require teachers to teach children how to use mathematics to solve problems, as is required in higher-level mathematics and many careers. For example, students may be presented with more word problems or "non-routine problems" in which they will need to think at higher levels and draw on their conceptual understanding of mathematics in order to discern the most efficient method of solving it before applying an algorithm to find the correct answer. Students will be expected to not only know how to find the answer to a problem but also possess a more refined conceptual understanding of its solution.

Ultimately, the standards' ability to enhance US students' mathematics achievement hinges on its implementation by teachers in classrooms (Cobb and Jackson 2011). Thus, effective implementation is dependent upon adequate professional development for educators, including evidence-based guidance on how to provide struggling students with appropriate additional supports and well-aligned curricula that recognize the diversity of learners (Cobb and Jackson 2011; Lee 2011; Powell et al. 2013). For example, Ginsburg et al. (2008) stress the importance of more relevant and rigorous content in preservice teacher training, including courses that introduce teachers in training to new research on children's mathematical thinking, facilitate their understanding and future use of formative assessment, teach them the basic mathematical ideas that undergird early childhood mathematics, expose them to a variety of strong curricula and appropriate pedagogy, teach them methods of assessment along how assessment should guide instruction, and have them analyze videos of exemplars teaching mathematics (Ginsburg et al. 2006; Ginsburg et al. 2008). Furthermore, Ginsburg and colleagues recommend "extensive, frequent, and long-term" in-service training that adheres to the same model and includes training in how to implement the required curriculum and critically reflect on one's practice. While the complexities of mathematics pedagogy are beyond the scope of this chapter, in general the CCSS in mathematics require students to develop a deeper level of both skill fluency and conceptual understanding than was previously found in many American classrooms. For a more in-depth treatment of best practices in mathematics pedagogy see Ginsburg et al. 2008; Powell et al. 2013; Russell 2012; VanDerHeyden and Alsopp in press.

The CCSS in mathematics seem to provide educational professionals with critical evidence-based direction toward increasing the mathematics proficiency of students in the USA, although in the short run, the pedagogical shift may create difficulties. Some older students may struggle since they will be expected to meet the new mathematics standards without the attendant foundation necessary for success (Porter et al. 2011; Powell et al. 2013). Moreover, students already experiencing mathematics difficulties will likely experience even greater challenges when faced with more difficult standards, especially since the CCSS guidelines offer minimal guidance on accommodating students with disabilities (Haager and Vaughn 2013; Powell et al. 2013; see for instance What is Not Covered by the Standards, NGACBP/CCSSO 2010). To best support students struggling to achieve the CCSS, Powell et al. (2013) recommend teachers focus on developing foundational skills to targeted standards (see Powell et al. 2013 for further guidance on working on foundational skills to help students access targeted *common core* standards). In addition, and as will be discussed in greater detail later, teachers' awareness and utilization of RTI practices in a multi-tiered system of support may serve as the linchpin that enables all students to achieve mastery of the content specified in the CCSS.

The ultimate success of the CCSS in mathematics is debatable. Cobb and Jackson (2011) point out that the developers' attention to research on learning progressions make the CCSS more coherent than many preexisting state standards. However, other scholars doubt the CCSS will do much to enhance the quality of mathematics instruction in American schools (Lee 2011; Porter et al. 2011). For instance, despite claims to the contrary, one study analyzing the CCSS in relation to previous state standards concluded that the CCSS are only slightly more focused in mathematics and less focused in ELA (Porter et al. 2011). In addition, the *Benchmarks* document focuses almost exclusively on the economic and political reasons the USA should align its standards with international benchmarks in mathematics and language arts. However, this document neglects to provide evidence to support why the international benchmark standards are optimal beyond the fact that similar national standards have been adopted by countries, often quite different from the USA, whose students currently outperform US students. The developers of the CCSS in mathematics claim that the new standards are modeled after the trends in international mathematics and science study (TIMSS) and mathematics standards in countries that consistently outperform US students on mathematics assessments (e.g., Singapore, China, South Korea, Japan, and Finland)(NGACBP/CCSSO 2010). However, Porter et al. (2011) found the content standards for Finland, Japan, and Singapore actually align with the CCSS at low levels (0.21, 0.17, and 0.13, respectively) due to those countries' greater emphasis on the performance of procedures over cognitive demand.

Standards for Other Academic Subjects

At this point, content standards have only been developed for literacy and mathematics. Content standards are, however, being developed for other academic subjects, including history, science, world languages, and arts. Although the CCSSO is not spearheading these efforts, they are similar to the standards created in mathematics and literacy since these standards will aim to hold all students to a high standard of achievement.

Teacher Education and Professional Development

In order for the shift to the new standards to be successful, teachers (both preservice and in-service) will need appropriate preparation in both the new standards and the updated pedagogy they will be expected to implement. The rigor and elevated demands of the CCSS require educators to both enhance their understanding and expand their repertoire of skills. At present, both teacher preparation and professional development are being managed at the local level and, as a result, vary from district to district and state to state. Some districts are providing professional development to all teachers on site, while other districts are only offering professional development to some teachers or utilizing free online professional development modules. The CCSS initiative has spurred several national organizations to put forth policy statements encouraging states to bring teacher education, professional-development programs, and teacher certification and licensure in line with the CCSS (King and Jones 2012). This could negatively affect teacher support of the CCSS (Ediger 2011; Lee 2011).

Consequently, thoughtful and thorough teacher preparation is a critical key to ensuring success under the new standards (Cobb and Jackson 2011).

Assessment

With new standards, a revamped assessment process has also been envisioned. Currently, two assessment consortia, the *partnership for assessment of readiness for college and career* (PARCC) (http://www.parcconline.org/) and the Smarter Balanced Assessment Consortium (SBAC) (http://www.smarterbalanced.org/), are each working with 24 states to develop new assessments aligned with CCSS. The development of these assessments represents a significant investment, as these two consortia have received a total of US$ 330 million through a grant from the US Department of Education to develop new assessment systems aligned to the standards (Alliance for Excellent Education 2013). The consortia are developing both formative assessments to track student progress during the year and end-of-year summative assessments. In theory, according to both SBAC and PARCC, benchmark assessments of students' skills will be administered at the beginning of the year, with the option of ongoing progress monitoring over the course of the year, culminating in a final summative assessment near the conclusion of the school year. Specifically, PARCC, articulates plans about offering nonsummative diagnostic materials that can be used mid-year and to track student progress (PARCC 2013). They will also be offering what they describe as "innovative assessment tools" for students in grades K and 1 that are supposed to aid teachers by giving them information about students' mastery of particular standards. Details of these assessments are not yet available nor is a time frame known. PARCC anticipates that its mid-year assessments will be available beginning in 2015. Smarter Balanced has released a *master work plan* for the formative assessments that they plan on offering (Smarter Balanced 2012). This document details the anticipated outcomes, parameters, and an estimated completion date of 2014. While the tools are described as agile and able to incorporate feedback from teachers and consortium member representatives, details of how this will take place is not immediately clear (Smarter Balanced 2012). Presently, it appears that each may include a mid-year assessment option; however, neither assessment system appears to provide an infrastructure that specifically facilitates frequent progress monitoring of students, despite claims to the contrary (PARCC 2013; Smarter Balanced 2012). Moreover, whereas both assessment consortia are projected to have a similar single administration cost per pupil compared to current state assessments, it is not clear what the source of funding may be to support the repeated administration of these assessments for screening, mid-year formative assessment, and progress monitoring.

Additional formative assessments are also being developed by other groups. For example, *Renaissance Learning* has made extensive efforts to align the *star reading and mathematics* assessments with CCSS. In school districts that have sufficient funding, teachers may utilize effective progress monitoring tools such as *star reading and mathematics* from *Renaissance Learning* to track their students' progress toward meeting year-end goals and adjust instruction accordingly. In addition, the mathematics assessment project (MAP), a collaboration between the Shell Center team at the University of Nottingham and the University of California, Berkeley, is developing formative assessments aligned with the CCSS. Grounded in research detailing the critical importance of formative assessment reviewed by Black and William (1998), the MAP formative assessments are of two kinds: *concept development lessons and problem-solving lessons*. Both types of formative assessments are rooted in the content described in the standards. The developers recommend teachers use the concept development lessons every few weeks to gauge and improve students' level of understanding; whereas the problem-solving lessons are designed to be utilized less frequently and are aimed at helping students connect all the mathematics they have learned and deepen their understanding of mathematics. The new assessments aim to move away from the simple regurgitation of information in multiple-choice formats, which are currently a common form of state assessment, toward tests that include performance tasks, which rely on production-type responding (NGACBP/CCSSO 2010). Due to the high cost of scoring the performance tasks and added difficulty in establishing and meeting high standards of scoring agreement, most states have drastically reduced the number of performance tasks included in summative assessments of skills.

Despite these aims, critics point out that the new assessments still rely heavily on multiple-choice and fill-in-the-blank formats, although the former have been repackaged as "selected response" questions, and that states have drastically reduced the amount of time originally allocated for the performance tasks (Applebee 2013). Critics further caution against the risk that new assessments, which they argue do not adequately reflect the standards, will come to drive instruction in lieu of the standards since the assessment "tasks that are being developed are tied to the CCSS, not to the curriculum" (Applebee 2013, p. 30).

The new assessments associated with CCSS will be computer-administered, intended to provide faster turnaround times for results, and increase the likelihood that teachers will utilize the assessment results to provide instruction more tailored to the needs of their students. In addition, the *smarter balanced* assessment system will utilize computer adaptive testing (CAT), which adjusts the difficulty of questions throughout the assessment based on accuracy of student responses. The CAT approach identifies students' proficiency more efficiently and is believed to provide more precise scores for most test-takers, especially those with extremely high or low scores (Thissen and Mislevy 2000; Weiss and Kingsbury 1984). In contrast to the computer adaptive summative assessment offered by SBAC, the PARCC system will offer a computer-based fixed-form summative assessment, a key difference between the assessments developed by the two consortia.

In theory, the CCSS and accompanying changes to instruction and assessment, with their widely touted research- and evidence base and increased rigor, hold great promise for K–12 education. Fulfilling that promise, however, means increasing the academic proficiency of *all* students across the full range of the achievement continuum to the level of international benchmarks. Naturally, some students will require additional support and academic interventions to perform at this level. Importantly, standards do not teach, teachers teach. Thus, it is not the standards in and of themselves that will foster the desired improvements in students' academic achievement. Rather it is teachers using evidence-based instructional strategies in the classroom each day that will actualize the CCSS outcomes (or mastery of CCS standards). Consequently, it is important for education professionals to consider how the CCSS will intersect with existing or developing RTI/MTSS frameworks.

Response to Intervention/Multi-Tiered Systems of Support

RTI consists of incrementally more intensive instruction provided based on student need as estimated by screening and progress monitoring data (Jimerson et al. 2007). Deno et al. (in press) highlight that RTI is founded on a problem-solving model aimed at promoting student achievement through the use of data-based decision-making. The premise of RTI is that most students should be able to succeed given core instruction in general education using high-quality, research-based instruction for all students. Indeed, under an appropriately implemented RTI program, 80% of students should have their needs met within the general education setting (Lawrence 2004; Batsche et al 2005; O'Conner 2003). For those students who are not successful in core instruction, there are increasingly intensive interventions that can be implemented. However, prior to beginning more intensive intervention, two criteria must be met. First, students must have access to a high-quality core curriculum and program of instruction (McMaster and Fuchs in press). Otherwise, the idea that most students will be able to learn in their general education classroom will not be the reality, disrupting the functioning of the whole system. Second, there must be a good way for teachers and schools to determine student risk or need for more intensive instruction. Through the use of data, RTI allows school professionals to verify student learning in response to particular levels of support or intervention and respond accordingly (decrease intervention intensity, increase intervention intensity; Barnett et al. 2004).

The challenge to using RTI in the intended manner is ensuring consistent use of high-quality instruction, data collection, and decision-making. When schools are not able to offer high-quality instruction or monitor student progress, then a school's resources may be drained. The CCSS provide an outlet for helping make RTI successful through working toward a high-quality educational foundation for all students while also presenting the potential for easier student data tracking.

The Shared Functions of CCSS and RTI

The CCSS represent the contemporary context of the American public education and RTI may offer a map to guide this transition. This shift to the CCSS provides education professionals an opportunity to reconsider curriculum and assessment practices, and simultaneously begin to envision how to create or use an RTI framework to help meet the new curricular guidelines. While individual schools and districts will have to evaluate the RTI structures that they have in place, the time is ripe for reconsidering what both RTI and CCSS can offer that will strengthen school instructional and intervention programs. The following section provides a brief discussion of areas of intersect between CCSS and RTI. What is included here is not exhaustive because as both CCSS and RTI continue to develop, it is anticipated that additional areas of intersect will continue to emerge. Nonetheless, education professional are encouraged to consider the relationship between CCSS and RTI as they implement CCSS and work to promote student success.

In order to successfully integrate CCSS and RTI, a few considerations need to be made. The first is school support. This includes having school staff, both administrators and teachers, help build and maintain the CCSS and RTI within their school contexts. While widespread involvement will automatically be incorporated into the adoption of CCSS by schools, since teachers and administrators need to familiarize themselves with the new guidelines as they are implemented, widespread involvement should also be part of either the adoption or enhancement of and RTI framework. In order to help ensure that this happens, the entire school staff should be involved in the initial organization and implementation of RTI. Oftentimes, this is done through a series of committees (e.g., needs assessment committee, reading intervention committee) to determine the best way to carry out specific parts of the program. Also part of this structure should be a problem-solving team. The task of this particular group is to meet regularly and look at student data at the individual, class, grade, and school-wide level. This team is often most successful if a variety of individuals, including teachers (general and special education), specialists, administrators, and parents, are active members of this team. The intervention data that they monitor should be used to guide interventions. A fundamental component of RTI is using sound methods to measure students' initial skills and progress in response to instruction. Finally, instruction needs to be organized in a tiered system with increasingly intensive interventions used for students who are not attaining mastery at lower tiers of instruction (Lembke et al. 2010). Thus, the infrastructure supporting RTI efforts in a school should also be valuable in facilitating appropriate screening and progress monitoring assessments, as well as encouraging high-quality instruction that targets growth for all students (whether they be in special education, general education, or gifted and talented programs).

Implications for Both General and Special Education

The RTI framework aims to allow as many students as possible to be successful within core instruction. However, for those students who are still struggling even with the implementation of increasingly intensive interventions, special education can provide additional resources to accomplish or sustain the most intensive levels of intervention in a system. In an RTI context, eligibility decisions are driven by data gathered about a student's performance over time and as more and more intervention services are put into place. As a result, students are able to receive an education in the least restrictive environment while still working to meet their educational needs. However, accurate decision-making requires the regular administration of valid and reliable assessments to guide and evaluate instructional effects. The implementation of CCSS does not include an emphasis on data collection and "there are no data collection requirements of states adopting the CCSS" (NGACBP/CCSSO 2010), although certainly there are implications for data collection that stem from CCSS.

According to the Individuals with Disabilities Education Act (IDEA), the application of the grade-level expectations is also relevant for students with disabilities. The expectation remains that students with disabilities will receive the supports and accommodations (e.g., *individualized education plans,* qualified teachers, and specialized instructional support personnel) necessary to enable them to meet the new standards. With the understanding that the CCSS will likely be more rigorous than the previous academic content standards in many states, it will be increasingly important that education professionals attend to the needs of all students; thus, an RTI infrastructure can be helpful to facilitate student success. It is possible that the appropriate assessment, research-based instructional practices, and progress monitoring associated with RTI and the tighter focus on key content associated with CCSS will increase the likelihood that students with disabilities are able to effectively access the curriculum and meet the new standards. CCSS materials do make mention of additional supports and services that some students with disabilities may be eligible to receive, such as the presentation of information in multiple formats, variety in assignments and educational products, and changes in materials and procedures (NGACBP/CCSSO 2010).

Considerations and Recommendations

The CCSS provide a direction for schools to increase their rigor without providing explicit steps toward meeting this end. As such, CCSS represents an opportunity to leverage or implement an RTI approach because it helps provide a framework to realize CCSS. With this in mind, there are endless possible ways to carry this out and this paper outlines a few considerations. To the degree that the CCSS emphasizes data-focused and quality instruction, this will likely make it easier for a school to assess and track student strengths and weaknesses, which is essential to a good RTI system. While every school will need to look at their individual goals and needs, general recommendations and considerations for how to use CCSS as an opportunity to begin RTI are described here (see Table 1 for a list of implications for practice).

Understanding the educational needs of a school population provides the structure for academic RTI programs. The shift to the CCSS further necessitates new curricula and assessment options (e.g., PARCC and SBAC) aligned with the CCSS (NGACBP/CCSSO 2010). While there is some controversy surrounding these kinds of standardized curricula and corresponding assessments, optimally these tools will be valuable for conducting school-wide needs assessments and establishing baseline performance for all students. Through the use of school-wide implementation of benchmark tools associated with these programs, schools are able to evaluate the specific needs of their student body. For instance, grade level reading and mathematics screenings used at the beginning of the school year may indicate appropriate levels of instruction for students relative to their achievement levels. Having this kind of systematic and school-wide data allows schools to target their RTI program to best use resources and serve students.

Appropriate assessment provides the opportunity to develop strong intervention programs. If appropriate assessments are developed for use with CCSS, then schools could examine their RTI system and determine instructional needs.

Table 1 Common core state standards and RTI/MTSS: implications for practice

General considerations when integrating CCSS and RTI
Implementation of the CCSS provides basic standards for schools
This transition to the CCSS offers an opportunity for educators to reconsider current curriculum and assessment practices
Utilizing evidence-based practices and data-based decision-making will enable educators to foster students' academic achievement within both the CCSS and RTI
Teacher preparation and professional development
The rigor and elevated demands of the CCSS require educators to expand their repertoire of skills
Thoughtful and thorough teacher preparation in both the new standards and the updated pedagogy is a critical key to ensuring success under the new standards
Assessment
Formative assessment and frequent progress monitoring is necessary to increase the academic proficiency of all students across the full range of the achievement continuum to the level of international benchmarks
Since neither assessment system associated with the CCSS appears to provide an adequate infrastructure for frequent progress monitoring of students, it is recommended that educators employ formative assessments being developed by other groups, such as *Renaissance Learning*, aimsweb, and MAP
Optimally, the new assessments will be valuable for conducting school-wide needs assessments and establishing baseline performance for all students. Schools should evaluate the specific needs of students and target instruction and interventions to their achievement levels based on collected data
Knowledge of assessment processes, purposes, and interpretation among teachers and administrators is essential
Data management infrastructure is important to document student achievement and growth across the grades
The intersect of CCSS and RTI
Successful integration of CCSS and RTI is dependent upon adequate school support, an active problem-solving team, close monitoring of student achievement and intervention data, data-based decision-making, and the organization of instruction in a tiered system of increasingly intensive interventions
An RTI framework can support schools as they strive to achieve the CCSS by conducting screening and progress monitoring assessments and supporting the provision of high-quality, evidence-based instruction
Emphasis should be placed on data-focused, quality instruction aligned with the CCSS and the regular assessment and tracking of student strengths and weaknesses, an essential component of RTI
The CCSS's overemphasis on summative assessment can be addressed by multi-tiered systems of support, such as RTI, wherein, teachers regularly employ sound formative assessments to monitor student progress
To optimize student outcomes, establish a system (e.g., a problem-solving team) for regularly evaluating student placements based on data and promptly make necessary changes
Implications for both general and special education
In an RTI context, educators make data-driven decisions based on information gathered via the regular administration of valid and reliable assessments, not just teachers' assessments. Educators utilize progress-monitoring data to track students' progress and guide instruction
Appropriate assessment, research-based instructional practices and interventions, and progress monitoring associated with RTI are critical tools to enable students with disabilities' to effectively access the curriculum and meet the new standards
Utilizing CCSS to reconsider or implement an rti framework
Implementation efforts must be systematic
Evidence-based practices and regular progress monitoring must be incorporated into the classroom to enable all students to make sufficient progress toward mastering academic material
Ultimately, implementing an RTI program that provides clear guidelines about student placement, staff participation, appropriate assessment, and intervention options to create a stronger learning environment for all students

CCSS common core state standards, *RTI* response to intervention

Such assessments would be advantageous to both students and schools. Through systematic assessment of academic weaknesses and needs, schools are able to identify appropriate instruction to target knowledge development and promote mastery learning to help students attain concepts and topics that are more challenging. Wang and Holcombe (2010) found that promoting positive student identification and providing positive and improvement-based praise, with an emphasis on effort and mastery, resulted in higher levels of motivation and achievement. Using CCSS to promote and improve RTI, offers schools a chance to shift their emphasis to ensure that it aligns with these goals.

In order to successfully use CCSS to reconsider or implement an RTI framework, it is important to be systematic in the implementation efforts. Manley and Hawkins (2013) provide guidance about how to do this. For instance, they describe how a teacher can track the progress of an entire class on a single Excel spreadsheet. There are assessments such as those available through *Renaissance Learning* and *Aimsweb* that work to provide an infrastructure to establish an understanding of student knowledge and also provide a means of monitoring student progress over time. Aimsweb offers a range of tools that allow schools to both set student goals and monitor student data. Their computer interface offers schools the chance to easily produce data about students' performance in ELA, mathematics, and reading. They have also announced new products including mathematics formative assessments that are aligned with the CCSS. Additionally, new reading reports are meant to offer information about anticipated student performance on CCSS reading materials (Aimsweb 2012). Similarly, Renaissance Learning (2013) offers assessment tools, aligned with CCSS, that are designed to be used for universal screening three times per year to help effectively gather and track student data. This is advantageous because it provides an opportunity to identify areas of proficiency and weakness at both the individual and aggregate levels. Once these patterns of performance have been established, it opens up the opportunity to allocate resources according to greatest need. For instance, Manley and Hawkins (2013) describe how teachers may share resources (e.g., time, materials) to help ensure more students in need get the necessary help. On a school-wide level, this could include strategies such as ensuring that students in specific grade levels receive additional time with a reading specialist or offering an after-school mathematics club to help with homework. While the options that schools have for interventions vary depending on infrastructure and resources, a prescribed assessment system allows schools to use the resources they have in the most efficient manner, by knowing student's academic needs and providing supports that are informed by empirical evidence.

Implementing the CCSS may help education professionals to ensure that their interventions are targeted to meet student needs, and may also help schools with a number of important logistical challenges. For example, when students transfer schools or even move states, schools may have a better sense of the student's present academic levels and any relevant strengths and weaknesses. This means that the school will be able to place the student in an appropriate instructional environment and offer interventions, all without a lengthy waiting period to examine the student's present academic performance level. Currently, schools are able to do this with students who have special accommodations, such as IEP or 504 plans, but not for other students. Shared standards, shared curricula, and assessment data will allow schools to more fluidly incorporate students into their school and ensure appropriate instructional level more readily.

An important consideration of CCSS when thinking about RTI is the emphasis on curriculum content mastery (Manley and Hawkins 2013). The CCSS emphasize the importance of not just teaching and learning, but aims to ensure that students master the material that is presented (Manley and Hawkins 2013). Although formative assessments are being developed by the two consortia, consistent with the CCSS's emphasis on mastery, the CCSS assessment system presently appears to focus almost exclusively on summative assessments of students' skills. By contrast, RTI requires teachers to employ sound formative assessments to monitor student progress in conjunction with summative evaluations of mastery. That is, effective RTI implementation requires that teachers track what students have learned, what they are still struggling with, and which individuals may be struggling relative to their classmates. Clearly, this is an area where multi-tiered systems of support, such as RTI, can help provide processes to support student success. To the extent that CCSS articulates a new destination for students, RTI may provide a means of getting there. For example, Manley and Hawkins (2013) describe how, in an RTI model, a teacher can align his or her teaching to the academic content standards and then use progress monitoring tools and formative assessments to determine exactly what students do and do not know. A teacher can then adjust his or her teaching to ensure that the students are able to master all of the material. This type of progress monitoring approach to assessment can reveal which students are keeping up with the class and which students are not mastering the material along with their classmates and indicate which students may need more intensive interventions. Progress monitoring also allows teachers to have a system for checking for student understanding (a tenet of effective instruction Hattie 2013) and for identifying performance patterns that signal the need for reteaching class wide, in small groups, or with individual children. Assessment is intended to be a feedback loop to the teacher to know if the teaching has worked or not. A strong RTI program in conjunction with a system of common core standards, such as the CCSS, could help ensure that general education instruction is effective for the majority of students in a class (Lembke et al. 2010).

Concerns and Critiques when Integrating RTI and CCSS

Critics of RTI have articulated that it can be difficult to implement as intended. At the core of RTI are a series of assumptions: (a) RTI allows professionals to evaluate practical proficiency, (b) the use of RTI provides continuous optimization of interventions, (c) RTI allows for better eligibility decisions, (d) RTI can promote achievement, and (e) RTI provides more flexibility for intervention services (VanDerHeyden et al. 2005). The common thread through each of these is finding a balance between structure and flexibility. Schools need to have systems in place that provide a framework for this system, yet, at its core it needs to provide flexibility to adjust interventions and student placement based on data. Perhaps one of the greatest strengths of CCSS is the chance to have consistency across schools, districts, and states while allowing for the room to have a system that can change and adapt based on individual and contextual needs.

Beginning to put RTI into place is initially a tremendous undertaking. It requires a lot of preparation in the beginning to secure staff buy in, create infrastructure, and reconsider how to best use school resources. The revision of curriculum and school organization associated with the CCSS affords an opportunity to begin to use or improve a school's RTI program. Ultimately, implementing an RTI program that provides clear guidelines about student placement, staff participation, appropriate assessment, and intervention options will create a stronger learning environment for all students.

Areas for Future Research

Evidence-based practices are essential in the RTI framework. While there is a research base for CCSS, further research is warranted to determine student outcomes associated with implementation of the CCSS. As the CCSS are implemented on a large scale over the next few years, additional data need to be collected. In particular, research should focus on how well students are able to meet the CCSS benchmark standards (including students with disabilities) and how well teachers are able to promote student achievement of the CCSS. Additionally, as schools and teachers have a chance to become accustomed to using the CCSS framework, it will be important to continue to monitor the percentage of students served within general education and special education classrooms, as this has important implications for the success and future of RTI in the schools. Specifically, future research should continue to look at the ways in which the CCSS and RTI are compatible and any changes that may be necessary for these goals to be met simultaneously. Given the importance of international benchmarks for the development of the standards, it will be important to examine and evaluate the extent to which the CCSS may impact student achievement over time relative to those benchmarks. The limited scope of the CCSS (reading and mathematics), overlook other important skills that will likely be important to facilitate the success of students in an increasingly global context (Wagner 2010). Furthermore, researchers should investigate whether or not the CCSS do in fact prepare students for college and career readiness, as they purport to do.

Conclusion

The intent of both RTI and CCSS is to further promote student achievement in the American education system. While there will be a period during which schools shift to the new curriculum standards, this transition time is a tremendous opportunity for schools to explore refining or developing an RTI infrastructure. Perhaps most importantly, education professionals are encouraged to explore how the implementation of CCSS and RTI will have reciprocal advantages for facilitating student achievement. Since both the CCSS and RTI philosophies are based on the idea of high-quality education programs and data-based decisions, the contemporary context affords an opportunity for education professionals to re-evaluate how to strengthen their RTI programs as they consider how to implement the CCSS.

References

Aimsweb. (2012). http://www.aimsweb.com/. Accessed 6 June 2013.

Alliance for Excellent Education. (2013). Common core state standards 101. http://all4ed.org/reports-factsheets/common-core-state-standards-101/. Accessed 7 July 2013.

Applebee, A. (2013). Common core state standards: The promise and the peril in a national palimpsest. *English Journal, 103*(1), 25–33.

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal, 100*(6), 14–27.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *The Journal of Special Education, 38,* 66–79.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., Reschly, D. J., Schrag, J., & Tilly III, W. D. (2005). *Response to intervention*. Alexandria: National Association of State Directors of Special Education, Inc.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*(2), 163–197.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education, 5*(1), 7–74.

Bomer, R., & Maloch, B. (2011). Relating policy to research and practice: The common core standards. *Language Arts, 89*(1), 38–43.

Calkins, L., Ehrenworth, M., & Lehman, C. (2012). *Pathways to the common core: Accelerating achievement*. Portsmouth: Heinemann..

Cobb, P., & Jackson, K. (2011). Assessing the quality of the common core state standards for mathematics. *Educational Researcher, 40*(4), 183–185.

Coleman, D. (2011). Using the common core state standards to drive student achievement and school improvement. Teach for America's 20th Anniversary Summit. http://www.teachforamerica.org/corps-member-and-alumni-resources/alumni-summits/20th-anniversary-summit. Accessed 6 June 2013.

Common Core State Standards Initiative. (2010). Development Process. Downloaded 7.7.2014. http://www.corestandards.org/about-the-standards/development-process/.

Confrey, J., Maloney, A. P., & Nguyen, K. (2010). *Learning trajectories display of the common core standards for mathematics*. New York: Wireless Generation.

Cuban, L. (2010, July 25). Common core standards: Hardly an evidence-based policy. http://larrycuban.wordpress.com/2010/07/25/common-core-standards-hardly-an-evidence-based-policy/. Accessed 7 July 2013.

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology, 100*(4), 907.

Deno, S. L., Burns, M. K., & Jimerson, S. R. (in press). *Data-based decision-making. Handbook on RTI: The science and practice of multi-tiered systems of support* (2nd ed.). London: Springer.

Early, D. M, Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., Crawford, G., Weaver, W., Howes, C., Ritchie, S., Kraft-Sayre, M., Pianta, R., & Barnett, W. S. (2005). Pre-kindergarten in 11 states: NCEDL's multi-state study of pre-kindergarten and study of statewide early education programs (SWEEP): Preliminary descriptive report. NCEDL Working Paper. http://www.fpg.unc.edu/node/4654. Accessed 6 June 2013.

Ediger, M. (2011). Assisting pupils in mathematics achievement (The common core standards). *Journal of Instructional Psychology, 38*(3), 154–156.

Foorman, B. (2012). Finding the science of reading in the common core state standards [PowerPoint Slides]. http://fcrr.org/finding_sci_reading_ccss_ibida_20121011.pdf. Accessed 7 July 2013.

Gardner, D. P., Larsen, Y. W., & Baker, W. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: US Government Printing Office.

Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4–6: A national survey. *Elementary School Journal, 110*(4), 494–518.

Ginsburg, H. P., Kaplan, R. G., Cannon, J., Cordero, M. I., Eisenband, J. G., Galanter, M., & Morgenlander, M. (2006). Helping early childhood educators to teach mathematics. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 171–202). Baltimore: Brookes.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. Social Policy Report. *Society for Research in Child Development, 22*(1).

Graham, S., & Harris, K. R. (2013). Common core state standards, writing, and students with LD: Recommendations. *Learning Disabilities Research & Practice,* New York: Guilford Press, *28*(1), 28–37.

Graham, S., MacArthur, C. A., & Fitzgerald, J. (Eds.). (2013). *Best practices in writing instruction*. Guilford Press.

Haager, D., & Vaughn, S. (2013). Common core state standards and students with learning disabilities: introduction to the special issue. *Learning Disabilities Research & Practice, 28*(1), 1–4.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668.

Hattie, J. (2013). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. New York: Routledge.

Hausken, E. G., & Rathbun, A. (2004). Mathematics instruction in kindergarten: Classroom practices and outcomes. In *American Educational Research Association meeting*. San Diego.

Jago, C. (2009). A history of NAEP assessment frameworks. http://www.nagb.org/content/nagb/assets/documents/who-we-are/20-anniversary/jago-frameworks-formatted.pdf. Accessed 6 Oct 2013.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). Response to intervention at school: The science and practice of assessment and intervention. In *Handbook of Response to Intervention* (pp. 3–9). New York: Springer.

King, J. E., & Jones, A. (2012). The common core state standards: Closing the school-college gap. *Trusteeship, 20*(2), 16–21.

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology, 101*(1), 136.

Lawrence, K. S. (2004). Executive summary of the NRCLD symposium on responsiveness to intervention [Brochure]. National Research Center on Learning Disabilities.

Layzer, J. I. (1993). Observational study of early childhood programs. Final Report. Volume I: Life in Preschool.

Lee, J. O. (2011). Reach teachers now to ensure common core success. *Phi Delta Kappan, 92*(6), 42–44.

Lembke, E. S., Garman, C., Deno, S. L., & Stecker, P. M. (2010). One elementary school's implementation of response to intervention (RTI). *Reading & Writing Quarterly, 26,* 361–373. doi:10.1080/10573569.2010.500266.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah: Lawrence Erlbaum Associates.

Manley, R., & Hawkins, R. J. (2013). *Making the common core standards work*. Thousand Oaks: Sage Publications.

Musun-Miller, L., & Blevins-Knabe, B. (1998). Adults' beliefs about children and mathematics: How important is it and how do children learn about it? *Early Development and Parenting, 7*(4), 191–202.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston: NCTM.

National Council of Teachers of Mathematics. (2006). *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence*. Reston: NCTM.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston: NCTM.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common Core State Standards (Application of Common Core State Standards to students with disabilities). National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington, DC. http://www.corestandards.org/assets/application-to-students-with-disabilities.pdf. Accessed 6 June 2013.

National Governors Association, Council of Chief State School Officers, and Achieve. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: NGA, CCSSO, and Achieve, Inc.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics learning study committee, center for education, division of behavioral and social sciences and education*. Washington, DC: National Academy Press.

Newkirk, T. (2009). *Holding on to good ideas in a time of bad ones: Six literacy principles worth fighting for*. Portsmouth: Heinemann.

Newkirk, T. (2013). Speaking back to the common core. In *(postscript to) Holding on to good ideas in a time of bad ones: Six literacy practices worth fighting for (2009)* (pp. 1–7).

O'Connor. (2003). Tiers of intervention in kindergarten through third grade. Paper presented at the Response-to-Intervention Symposium, December 4–5, 2003, Kansas City. www.nrcld.org/html/symposium2003.

Partnership for Assessment of Readiness for College and Careers. (2013). Parconline.org.

Picklo, D. M., & Christenson, S. L. (2005). Alternatives to retention and social promotion: The availability of instructional options. *Remedial and Special Education, 26*(5), 258–268.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards the new US intended curriculum. *Educational Researcher, 40*(3), 103–116.

Powell, S. R., Fuchs, L. S., & Fuchs, D. (2013). Reaching the mountaintop: Addressing the common core standards in mathematics for students with mathematics difficulties. *Learning Disabilities Research & Practice, 28*(1), 38–48.

Renaissance Learning. (2013). Make a seamless transition to common core state standards. http://www.renlearn.com/ccss/default.aspx. Accessed 7 July 2013.

Russell, S. J. (2012). CCSSM: Keeping teaching and learning strong. *Teaching Children's Mathematics, 19*(1), 50–56.

Schmidt, W., Houang, R., & Cogan, L. (2002). A coherent curriculum. *American Educator*, *26*(2), 1–18.

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders, 32*(1), 7–18.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427–515.

Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer (Ed.), *Computerized adaptive testing: A primer*. Mahwah: Lawrence Erlbaum Associates.

Tienken, C. H. (2011). Common core standards: The emperor has no clothes, or evidence. *Kappa Delta Pi Record, 47*(2), 58–62.

VanDerHeyden, A. M., & Alsopp, D. (in press). Innovation configuration for mathematics.

VanDerHeyden, A. M., Witt, J. C., & Barnett, D. W. (2005). The emergence and possible futures of response to intervention. *Journal of Psychoeducational Assessment, 23,* 339–361.

Wagner, T. (2010). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need–and what we can do about it*. New York: Basic Books.

Wang, M. T., & Homcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Researcher, 47,* 633–662. doi:10.3102/0002831209361209.

Weiss, D. J., & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*(4), 361–375.

Wu, H. (1999). Basic skills versus conceptual understanding. *American Educator, 23*(3), 14–19.

Part III
Tier 1– Assessment, Problem Analysis, and Intervention

Screening Assessment Within a Multi-Tiered System of Support: Current Practices, Advances, and Next Steps

Nathan H. Clemens, Milena A. Keller-Margulis, Timothy Scholten and Myeongsun Yoon

The judicious, timely, and strategic delivery of targeted intervention strategies within a school system depends upon a reliable system of identification. Educators must have confidence that they can identify the students in need of supplemental intervention, and avoid identifying students not in need of such services. Thus, *screening assessment,* the strategies, tools, and processes used to identify individuals in need of supplemental supports, is considered one of the essential components of a response to intervention (RTI) system (Batsche et al. 2006; Gersten et al. 2009b). The importance of screening should not be understated. Intervening early, before early risk factors become intractable difficulties, offers the best hope for successful outcomes and prevention of long-term disabilities (Gersten et al. 2005; Torgesen 1998). Indeed, Fuchs and Vaughn (2012) suggested the "greatest accomplishment" (p. 196) of RTI may be the increased frequency with which schools are now implementing systems of early identification for students at risk of academic difficulties.

Although RTI and multi-tiered systems of support (MTSS) are relatively new terms, many of the processes they encompass, such as screening, are not. Screening for vision, hearing, and speech difficulties have been common in schools for some time, and history is replete with examples such as "readiness" testing for reading or mathematics difficulties at school entry (e.g., Leeand et al. 1934; Tower 1973). Screening assessment has also been used for screening elementary students for emotional/behavioral disorders or intellectual giftedness. But the advent and application of RTI/MTSS models has brought new attention to the use of screening in educational contexts, and terms such as "universal screening" or "benchmarking" have become part of the common vernacular in US schools. Screening has taken such an important position that it is required by law in some states as part of the process for determining special education eligibility (Zirkel and Thomas 2010).

Coinciding with the increased implementation of RTI models, screening assessment has been the focus of much research and scholarly activity across the last decade. Several excellent publications have described the important characteristics of screening tools, as well as skills recommended for assessment (Glover and Albers 2007; Gersten et al. 2009a, b; Jenkins et al. 2007). This work will not be reiterated. Instead, this chapter is meant to prompt innovation in our understanding of screening for academic difficulties. First, a review of common screening practices is provided.

N. H. Clemens (✉) · M. Yoon
Texas A&M University, College Station, USA
e-mail: nclemens@tamu.edu

M. A. Keller-Margulis
University of Houston, Houston, USA

T. Scholten
Ossining Union Free School District, New York, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_12

Screening: Observed Practices and Common Measures

Several recent studies have investigated the characteristics and practices observed in school-wide implementation of RTI models. Although most schools implement screening three times per year (fall, winter, spring), frequency varies widely, ranging from once per year to as often as five times per year, or even weekly (Prewett et al. 2012; Jenkins et al. 2013; Mellard et al. 2009; Tackett et al. 2009). The types of screening tools also vary; schools tend to use published tools more often (e.g., Dynamic Indicators of Basic Early Literacy Skills (DIBELS), AIMSweb), but other measures such as computer-adaptive assessments, materials that accompany curricula, and extant data (e.g., state tests) are also used. Reading and early literacy is the most frequently targeted area for screening, but screening for mathematics skills is also prominent, followed by written expression in fewer cases (Prewett et al. 2012). Many screening tools focus on skills known to be important indicators of subsequent development of skills proficiency in reading, mathematics, or written expression, informed by several expert panel reports and meta-analyses (e.g., Graham et al. 2012; National Reading Panel 2000; National Early Literacy Panel 2008; National Mathematics Advisory Panel 2008; RAND Mathematics Study Panel 2003; Snow et al. 1998).

Reading A good deal of research has focused on curriculum-based measurement (CBM) tools for use in universal screening applications for reading. CBM tools are attractive due to their simplicity, efficiency, ease of administration and interpretation, and suitability for screening decisions (Deno 2003). The assessment of oral reading (i.e., reading-CBM (R-CBM)) is the most commonly used reading screening assessment. R-CBM involves the individual administration of grade-level passages and the metric of interest is the number of words read correctly in 1 min. Research has established the validity of CBM-R as an index of overall reading achievement (Fuchs et al. 1988; Reschly et al. 2009; Wayman et al. 2007).

In middle to late elementary school and beyond, the value of R-CBM as an indicator of overall reading achievement diminishes (Marcotte and Hintze 2009; Shapiro et al. 2006; Yovanoff et al. 2005), likely due to factors such as the deceleration of fluency growth as students get older, the de-emphasis on decoding in instruction, and that as sufficient fluency is attained to support comprehension, additional gains in fluency are not associated with commensurate gains in comprehension. Work has examined alternatives to R-CBM such as CBM Maze, a group-administered timed cloze task, as a screener at the upper elementary levels and beyond. Some have found that CBM Maze measures have adequate reliability and validity for use in screening (Graney et al. 2010), while others have suggested concerns regarding reliability when only one Maze score is used for screening (Mercer et al. 2012a), and that R-CBM is a better predictor of overall reading achievement than maze measures (Ardoin et al. 2004). Maze, when included with indicators of word-reading fluency and teacher evaluations, has demonstrated effectiveness as a screener for identifying reading problems of students in fourth grade (Speece et al. 2010).

Other group-administered measures of reading achievement exist including tests of silent reading fluency, which consist of sentence verification tasks (Wagner et al. 2009) or measures in which students identify words within random word strings (Mather et al. 2004) or in context (Hammill et al. 2006). Sentence verification shows distinct promise as a universal screening tool; Denton et al. (2011) found that a sentence verification measure demonstrated equivalent correlations to measures of reading comprehension as individually administered R-CBM, stronger correlations to comprehension than Maze, and similar or better classification accuracy than R-CBM or Maze in predicting outcomes on a state high-stakes assessment. Speece et al. (2010) also found that a group-administered measure of silent word-reading fluency contributed meaningfully to a set of group-administered predictors that demonstrated similar accuracy in predicting fourth-grade reading outcomes over individually administered measures. Group-administered measures are extremely attractive as screening options, and continued work is needed to establish their validity in screening contexts.

Group-administered multiple-choice tests of comprehension have been used as screening tools, and such measures developed using a CBM framework have shown initial promise for screening (Anderson et al. 2011). Other standardized, multiple-choice measures of reading comprehension that have been used as screeners include the Iowa Test of Basic Skills, the Gates-MacGinitie Reading Test, and the Stanford Achievement Test series. Measures such as these have their advantages in being group-administered and are focused specifically on reading comprehension skills; however, standardized tests can be expensive to purchase and score.

Early Literacy Universal screening in early literacy typically uses measures of developmental precursors to reading including alphabetic knowledge (e.g., accuracy or fluency in naming letters or letter sounds), phonemic awareness (e.g., blending or segmenting words into phonemes), and decoding or word recognition. The predictive power of alphabetic knowledge is well known (Foulin 2005; National Early Literacy Panel 2008); thus, measures of letter naming or letter–sound identification are often recommended as screening options for kindergarten (Fuchs et al. 2004), and measures of word-reading fluency (i.e., word identification fluency) have demonstrated promise as screening options in early first grade (Clemens et al. 2011; Compton et al. 2006; 2010; Speece et al. 2011; Zumeta et al. 2012). Several tools are available to educators including AIMSweb (NCS Pearson 2012), DIBELS (Good and Kaminski 2002), easyCBM (Alonzo et al. 2006), and iSTEEP; www.isteep.com). Other batteries, such as the *Texas Primary Reading Inventory* (TPRI 2013) and *mClass: CIRCLE* (Landry et al. 2007) measures provide screening options using multiple literacy indicators.

Mathematics Like reading and early literacy screening, mathematics screening can also take several forms (see Lembke et al. 2012). Most common in elementary grades are measures designed to assess mathematics computation skills, and in some cases, problem-solving and applications. Mathematics CBMs meet the basic criteria for a universal screening in that there is evidence of their technical adequacy (see Christ et al. 2008; Thurber et al. 2002), they are efficient (often group-administered), and effective for the purpose of screening (e.g. VanDerHeyden and Burns 2005). Additionally, mathematics CBMs are related to broader measures of mathematics competency such as standardized, statewide achievement tests. Measures of computation and concepts/applications appear to offer fairly comparable levels of classification accuracy for performance on comprehensive, statewide measures of mathematics performance (Keller-Margulis et al. 2008; Shapiro et al. 2006). Measures assessing skills in number operations, algebra, and geometry have been developed, and preliminary research indicates sufficient technical adequacy for screening (Anderson et al. 2010; Clarke et al., 2011).

Early Numeracy Similar to the ways in which phonological awareness and alphabetic knowledge are developmental precursors of reading, growing consensus acknowledges the importance of early numeracy and "number sense" as foundational skills for mathematics development. Measures of early numeracy typically include tasks of number recognition, oral counting, quantity discrimination, pattern recognition, and other skills. Current evidence suggests that measures of early numeracy demonstrate adequate technical properties including both concurrent and predictive validity when used at the kindergarten and first-grade levels (e.g. Martinez et al. 2009; Missal et al. 2012; Lee et al. 2012). Additional studies have found that these measures demonstrate adequate classification accuracy making them appropriate to consider for screening (e.g., Baglici et al. 2010; Jordan et al. 2010; Lembke and Foegen 2009; Locuniak and Jordan 2008; VanDerHeyden et al. 2011). Measures of quantity discrimination in particular have been noted as potentially powerful predictors of mathematics calculation skills and are an attractive option for universal screening (Seethaler and Fuchs 2010; Seethaler et al. 2012).

Written Expression The area of writing has received less attention as it relates to universal screening. Recently, however, there has been increased interest in written expression given its importance as an essential skill for school success and postgraduation opportunities (Graham and Perrin 2007). The existing research on universal screening in writing has been dominated by the study of written expression CBMs (WE-CBM). Traditional WE-CBM procedures involve scoring students' responses to a writing prompt or story starter using a variety of metrics including the total words written (TWW), correct word sequences (CWS), and words spelled correctly (WSC). Studies of the technical adequacy of WE-CBM have indicated adequate reliability and validity (e.g., Deno et al.1982; Parker and Tindal 1989; Gansle et al. 2006). Additional studies have examined their use as screening tools at the elementary and secondary levels (McMaster and Campbell 2008; McMaster and Espin 2007; McMaster et al. 2009; McMaster et al. 2012). Additional questions remain, however, regarding the utility of production-dependent metrics (e.g., TWW, CWS, WSC), production-independent metrics (e.g., *percentage* of CWS or WSC), or accuracy indices (e.g., correct minus incorrect writing sequences) across age groups and for males or females (Espin et al. 2000; Malecki and Jewell 2003; Jewell and Malecki 2005; Mercer et al. 2012b).

Studies have also explored variations of writing prompts that may be more appropriate for younger students, including the use of letter-copying tasks or dictation (Lembke et al. 2003), word and sentence writing (e.g. Coker and Ritchey 2010; McMaster et al. 2009), handwriting (Van Waelvelde et al. 2012), and spelling (Puranik and Al Otaiba 2012). Results of the studies of writing measures for use with younger students have included adequate reliability and validity as well as sensitivity to growth over time (McMaster et al. 2009) suggesting utility as universal screening measures, although additional research is necessary to more clearly identify ideal skills for writing screening.

Academic Screening: The Need for Improvement

The work accomplished to date has provided a good foundation for implementing academic screening assessment within a MTSS. However, there is much room for improvement. Across academic areas, research on academic skills screening reveals that screening measures are inconsistent in correctly identifying the students who truly go on to experience academic difficulties. In many cases, studies have identified problems with screening measures identifying too many students as "at risk" who pass the subsequent outcome assessment (i.e., false positives) and problems with overall classification accuracy (e.g., Fuchs and Vaughn 2012; Clemens et al. 2011; Goffreda and DiPerna 2010; Jenkins et al. 2007; Johnson et al. 2009; 2010; Riedel 2007; VanDerHeyden 2011), and floor effects associated with some measures that negatively impact their accuracy (Catts et al. 2009). In the following sections, possible reasons for this inconsistency are described and the ways to potentially improve screening assessment are discussed.

Screening Content One reason for inadequate outcomes from screening may pertain to the content of the screening tools. In some cases, the authors may be underrepresenting, or perhaps misrepresenting the skills important to assess as important precursors and indicators of subsequent outcomes. Several types of skills, behaviors, and individual characteristics may deserve more attention when considering universal screening.

Language-Related Skills Oral language and vocabulary skills are critical foundational skills for subsequent reading comprehension and written expression, and instruction in these areas is receiving increased emphasis in kindergarten and preschool (e.g., Coyne et al. 2010; Gonzalez et al. 2011). Jenkins et al. (2007) advocated for greater attention to skills such as expressive and receptive vocabulary, sentence imitation, and story recall as screening options to provide a more complete picture of literacy skills.

Vocabulary knowledge explains unique variance in the prediction of reading comprehension skills over and above that accounted for by reading fluency, particularly for older students (Yovanoff et al. 2005), and vocabulary skills are garnering attention as potentially important components of early screening batteries (Alonzo et al. 2012; Norwalk et al. 2012; Parker and Ditkowsky 2006). Additionally, assessment of spelling and early writing skills in preschool and kindergarten students is gathering evidence for informing early screening for reading and written language difficulties (e.g., Clemens et al. 2014; Masterson and Apel 2010; Puranik and Lonigan 2012).

Behavioral Regulation and Attention The links between attention difficulties, impulsivity, and externalizing behavior problems with later academic difficulties are well established (Hinshaw 1992), and students with co-occurring behavior and academic difficulties early in school are at the greatest risk for negative outcomes (Darney et al. 2013). Many of us have heard teachers lament "I think he could learn it if he could just sit still." Indeed, attention difficulties and problem behaviors have been shown to negatively affect responsiveness to instruction (Al Otaiba and Fuchs 2006; Torgesen 2000; Hagan-Burke et al. 2011). Preschool students' behaviors including attention, effortful control, and response inhibition have demonstrated links with subsequent academic achievement, and simple screening tasks exist to measure these skills (Allan and Lonigan 2011). Incorporating assessments of students' ability to sustain attention and self-regulate with assessments of academic skills may help improve our identification of students less likely to respond to high-quality interventions, and at-risk for long-term academic difficulties. Additionally, these assessments may help educators better plan and incorporate strategies for fostering stronger behavior regulation and learning-related skills, which may be as important for future academic outcomes as the academic skills themselves (McClelland et al. 2006).

Specific Cognitive Processes Recent research has also revealed the potential for the assessment of specific cognitive skills such as phonological processing, rapid automatized naming (RAN), processing speed, working memory, and nonverbal reasoning to inform the identification of students at risk for learning disabilities in reading and mathematics (Catts et al. 2012; Compton et al. 2012; Fuchs et al. 2005, 2006, 2012; Tolar et al. 2012). For example, Fuchs et al. (2012) demonstrated significant reductions in false-positive decision errors using a second-stage screening that included measures of RAN, phonological processing, and working memory. Working memory is implicated across academic skill areas, and assessments of working memory may improve our ability to accurately identify students at risk for negative academic outcomes. More research is needed, however, to assess the unique or independent predictive value that particular cognitive processes may provide. The key is to find the specific variables that can be measured quickly and inexpensively, but most importantly improve our accuracy in identifying at risk learners.

Reconsidering Screening Format

Schools rely heavily on screening measures that are individually administered on a 1:1 basis with students (Jenkins et al. 2013; Mellard et al 2009). Efforts to improve the ease and efficiency of screening should consider alternate methods of administration.

Computer-Administered Measures and Computer-Adaptive Tests Computer-administered measures are referred to as those that mirror paper/pencil individually administered measures, but that students complete independently on a computer. Computer-administered measures have existed in school settings for quite some time (e.g., Hasselbring 1984), as have computer-administered CBM measures (e.g., Fuchs and Fuchs 1992); however, more computer-administered options are now emerging across academic areas (www.EasyCBM.com; www.isteep.

com). Research on the comparability of scores between computer- and paper-based assessments has found an overall consensus that scores are comparable regardless of assessment modality (Wang et al. 2008). Longer administration times of some computer-based measures may be offset by reducing adult supervision needed for screening assessment, as well as automatic scoring and data reports.

In contrast, computer-adaptive tests (CATs) employ software that automatically adjusts the difficulty of items based on the students' success or failure on a previous item. They offer benefits beyond that of the computer-administered measures because of the targeted and efficient manner in which student skills are examined. Advances in technology have made CATs more affordable and practical, and several CAT options now exist for screening in early literacy, reading, and mathematics (Formative Assessment System for Teachers 2013; Northwest Evaluation Association 2013; Renaissance Learning 2010). Measures often include categorical summaries of students' achievement or proficiency based on their scores, from which risk status can be estimated.

Contrasts of paper and pencil and CAT measures for early literacy suggest that CAT measures may be more efficient or cost effective (McBride et al. 2010); however, much work remains to establish the validity and classification accuracy of CATs for screening decisions. Shapiro and Gebhardt (2012) found that a CAT measure of mathematics demonstrated limited ability to reliably identify at-risk learners and generally did not improve over paper-based CBM measures of mathematics. Other work with a CAT measure of early literacy found it to reliably identify kindergarten students on track for successful reading outcomes; however, it did not completely replace individually administered paper-based measures in accurately identifying at-risk readers (Clemens et al. in press).

Additional work must determine whether CATs are appropriate across all grade levels and academic skill areas. CATs are often designed for students to take individually without adult support, and although by middle elementary students might be expected to possess sufficient self-regulatory skills for these tasks, this may be less certain for students in kindergarten and first grade. For example, in a study of a computer-based intervention with 5-year-old students, Kegel et al. (2009) found that students' self-regulatory skills impacted their behavior on the computer, and observed higher rates of off-task behavior (e.g., random mouse activity) among lower-achieving students.

Teacher Ratings and Nominations Teacher ratings and nominations are an attractive option for screening, due to their potential ease and efficiency of administration; a teacher nomination form might be completed in a few minutes, in contrast to the amount of time needed carry out individual assessments with a classroom of students. Teacher nominations and ratings of students behavior and academic skills have a long history of use in assessment contexts (e.g., DiPerna and Elliott 1999) and are used extensively in screening tools to identify students at risk for emotional or behavioral difficulties (Feeney-Kettler et al. 2011; Lane et al. 2009).

Research on the validity of teacher ratings and judgments of academic skills is mixed. One reason for mixed findings may be due to the different aspects in which teachers are asked to rate. Although teacher judgments are often observed to be moderately correlated with students' reading and mathematics achievement, when asked to estimate students' scores on a particular assessment or task (e.g., estimate number of words a student may read per minute, or number of mathematical problems completed) accuracy is inconsistent. Teachers tend to overestimate their students' reading fluency (Begeny et al. 2008; Feinberg and Shapiro 2009; Hamilton and Shinn 2003; Meisinger et al. 2009) and may overestimate their students reading growth over time (Graney 2008). In mathematics, moderate correlations have been observed between teachers' judgments of elementary and preschool students' mathematics skills and with direct assessments (Eaves et al. 1994; Teisl et al. 2001; VanDerHeyden et al. 2006), but similar to reading, studies indicate low agreement when teachers are asked to estimate mathematics scores (Eckert et al. 2006; Kilday et al. 2012).

Conversely, teacher judgment measures appear to demonstrate greater utility and consistency when they are asked to rate students' overall academic skills or proficiency using a standardized measure (Eckert et al. 2013; Hecht and Greenfield 2001; Hoge and Coladarci 1989). Speece and Ritchey (2005) found that teachers' ratings of their first graders' overall academic competence accounted for unique variance in year-end reading skills when used with direct assessments of reading and cognitive skills. Hence, research on the use of teacher ratings and judgments appears promising when teacher ratings are combined with direct assessments. It is found that the combination of direct assessments and teacher ratings of early reading skills improved their identification of at-risk readers (although the combination also appeared to increase the number of false positives). Speece et al. (2011) included teacher ratings in a battery of screening assessments to identify first-grade students at risk for reading problems and found that an accurate and parsimonious screening model included teachers' ratings of reading problems with two direct measures of word-reading fluency. Similar results have been found in other studies, as teacher ratings combined with direct assessment measures led to improvements in accuracy over either type of assessment alone (Snowling et al. 2011; Speece et al. 2010).

In addition to their logistical advantages, teacher judgments may be of particular benefit when evaluating overall achievement of complex and multifaceted academic skills (Eckert et al. 2013). For example, reading comprehension is a broad construct that can be difficult to ascertain through direct assessments (see Keenan et al. 2008); however, teachers are privy to a host of important behaviors in their daily interactions with students that may inform evaluation of the text-processing abilities. Teachers observe their students' proficiency and sophistication in answering questions, summarizing text, receptive and expressive vocabulary, as well their motivation to read; all aspects that are difficult to capture via direct assessments but may inform reading evaluation. Scholars have advocated for the inclusion of teacher judgment measures to more comprehensively evaluate comprehension skills (Fletcher 2006). Additional work is needed, however, as some studies have indicated the tendency for lower validity of teacher ratings for students with lower skills (Cabell et al. 2009; Begeny et al. 2008; Begeny et al. 2011; Feinberg and Shapiro 2009; Demaray and Elliot 1998; Hoge and Coladarci 1989; Martin and Shapiro 2011) or with younger students (Kenny and Chekaluk 1993). Some teachers may be better raters than others (Hoge and Coladarci 1989), and length of time in which the teachers have known their students is likely to impact rating accuracy; thus, using ratings to screen very early in the school year may be problematic. Additionally, in the transition from elementary to middle and high school, the ratio of teachers to students increases dramatically, thus teachers at higher grade levels may have less knowledge of their students in order to accurately rate the performances of their students.

In summary, the research to date has provided a good starting point on the types of measures and content that make up screening models. However, much work is left to improve both the accuracy and efficiency (these terms need not be mutually exclusive) of screening assessment of academic skills. The authors next turn their attention to methods used to make decisions based on screening measures, and methods used to evaluate the accuracy of screening tools.

Determining Whom to Treat and Evaluating the Accuracy of Screening Measures

Determining which students to identify for supplemental interventions is often based on cut scores intended to differentiate between students considered to be at risk for failing to achieve a subsequent outcome (or identified with a disability), or students not considered to be at risk. In practice, schools rely on cut scores or "benchmarks" established by published measures (e.g., DIBELS), local norms, percentile distributions, or professional judgment (Mellard et al. 2009). Of these methods, cut scores have been recommended for

		Outcome Assessment Result	
		Positive (*Failed outcome assessment, or disability present*)	Negative (*Passed outcome assessment, or disability not present*)
Screening Test Result	Positive: (*"At-risk" on screen*)	A *True Positives*	B *False Positives*
	Negative (*Not "at-risk" on screen*)	C *False Negatives*	D *True Negatives*

True Positive	Cell A. Students whom the screening test indicated were at risk, and failed the outcome assessment or were identified with a disability.
True Negative	Cell D. Students whom the screening test indicated were not at risk, and passed the outcome assessment or were not identified with a disability.
False Positive	Cell B. Students whom the screening test indicated were at risk, but passed outcome assessment or were not identified with a disability.
False Negative	Cell C. Students whom the screening test indicated were not at risk, but failed outcome assessment or were identified with a disability.
Sensitivity	A/(A+C). Proportion of students who failed the outcome assessment or identified with a disability that were correctly identified as at-risk on the screen.
Specificity	D/(D+B). Proportion of students who passed the outcome assessment or were not identified with a disability that were correctly identified as not at-risk on the screen.
Positive Predictive Value	A/(A+B). Proportion of students identified as at-risk on the screen that failed the outcome assessment or were identified with a disability.
Negative Predictive Value	D/(D+C). Proportion of students identified as not at-risk on the screen that passed the outcome assessment or were not identified with a disability.
Accuracy	(A+D)/(A+B+C+D). Overall accuracy of the screening test in correctly classifying individuals according to their status on the outcome assessment.
Base Rate	(A+C)/(A+B+C+D)The overall prevalence of the disorder in the population (i.e., the percentage students in the sample that failed the outcome assessment or were identified with a disability)

Fig. 1 2×2 contingency table and classification accuracy indices considered within the context of academic skills screening

screening decision-making (Gersten et al. 2009a; Mellard et al. 2009).

Cut scores may be derived either by the publisher of a screening tool, researcher, or locally by a user. In the following section, the process commonly used to determine cut scores on a screening tool is reviewed, in addition to the related methods of evaluating the accuracy of screening tools in correctly identifying students in need of supplemental interventions.

Classification Accuracy In an educational screening context, classification accuracy refers to the consistency at which a screening measure correctly classifies students according to a subsequent outcome, such as students who later pass or fail a future assessment. Typically, a 2×2 contingency table (see Fig. 1) is used to determine the screening measure's accuracy in correctly classifying students with the problem of interest (i.e., "positives," students who perform below a criterion level on the subsequent outcomes measure) and students without the problem (i.e., "negatives," students who perform above a criterion level on the outcomes measure). An assumption is that inaccuracy of the screening measure results in misidentification, which may produce false-positive errors (students for whom the screening measure indicated they were at risk for failure, but who actually passed the outcome assessment), and false-negative errors (students for whom the screening measure indicated they were not at risk, but who failed the outcome assessment). Additional terms are defined in Fig. 1. To estimate classification accuracy, one must have a complete data set that contains both the screening prediction and the actual outcome for a group of students. Thus, classification accuracy metrics can be used to estimate the accuracy with which a screening measure can be used with similar children, under similar circumstances, to make a similar type of judgment.

Deriving Cut Scores Currently, receiver operating characteristic (ROC) curve analyses are a predominant method for deriving cut scores. ROC analyses originated from work in signal detection, which involved human observers making perceptual judgments on incoming stimuli. Early applications in communication and defense systems evaluated the accuracy with which individuals could detect the presence or absence of an incoming signal (Egan et al. 1956; Swets et al. 1961). From work in signal detection, ROC curves were applied across a variety of disciplines including psychology and medicine, with applications in radiology (Metz et al. 1973) and evaluating the accuracy of diagnoses (Goodenough 1975). ROC curves were subsequently applied to work in education (Harber 1981), and have since become one of the primary and most frequently used methodologies in educational screening for deriving cut scores and evaluating the accuracy of screening tools.

An ROC curve is a plot of the rate of true positives against the rate of false positives across a range of potential cut scores, thereby yielding the sensitivity and specificity associated with scores on the screening tool. Sensitivity and specificity share an inverse relationship, and this distribution can be used to select a cut score based on the users' preference to maximize sensitivity, specificity, or to balance the two. Since false-negative errors represent arguably the most problematic type of error (i.e., failing to identify a student for intervention that is truly in need), Jenkins et al. (2007) advocated for selecting cut scores that prioritize the identification of students who are truly at risk (and thus avoid denying intervention to students truly in need), and recommended a minimum sensitivity of 0.90 (i.e., a 10% false-negative rate). Ideal rates for specificity are less clear; some have noted that specificity of 0.50 (i.e., 50% false-positive rate) is acceptable since the progress monitoring and intervention processes associated with an RTI model are designed to subsequently reduce false-positive errors further (Catts et al. 2009). Others have aimed for higher levels of specificity (i.e., above 0.80, Compton et al. 2006; 2010). After a cut score is identified, a 2×2 contingency table can be completed in order to verify and evaluate the accuracy of the screening measure in correctly classifying students based on the cut score chosen.

Data-Based Example Analyses of a data set to help illustrate this use of classification accuracy and ROC curves to generate cut scores have been included. These data were drawn from an archival data set from a school district in the Northeast USA, which included three cohorts of kindergarten students followed through the end of first grade ($N=755$). Screening measures included letter-naming fluency, letter–sound fluency, and phoneme segmentation fluency administered in the winter of kindergarten. Reading outcomes were evaluated at the end of first grade using R-CBM, and to reflect a minimum standard for acceptable reading, the 25th percentile based on national normative data (40 words correct per minute) was selected. A total of 200 students scored below this criterion at the end of first grade, yielding a risk base rate of 26.5%.

ROC curves were generated for each of the three kindergarten measures (see Table 1), and the resulting ROC curves are displayed in Fig. 2. Letter–sound fluency (LSF) demonstrated the highest area under the curve (AUC; an index of

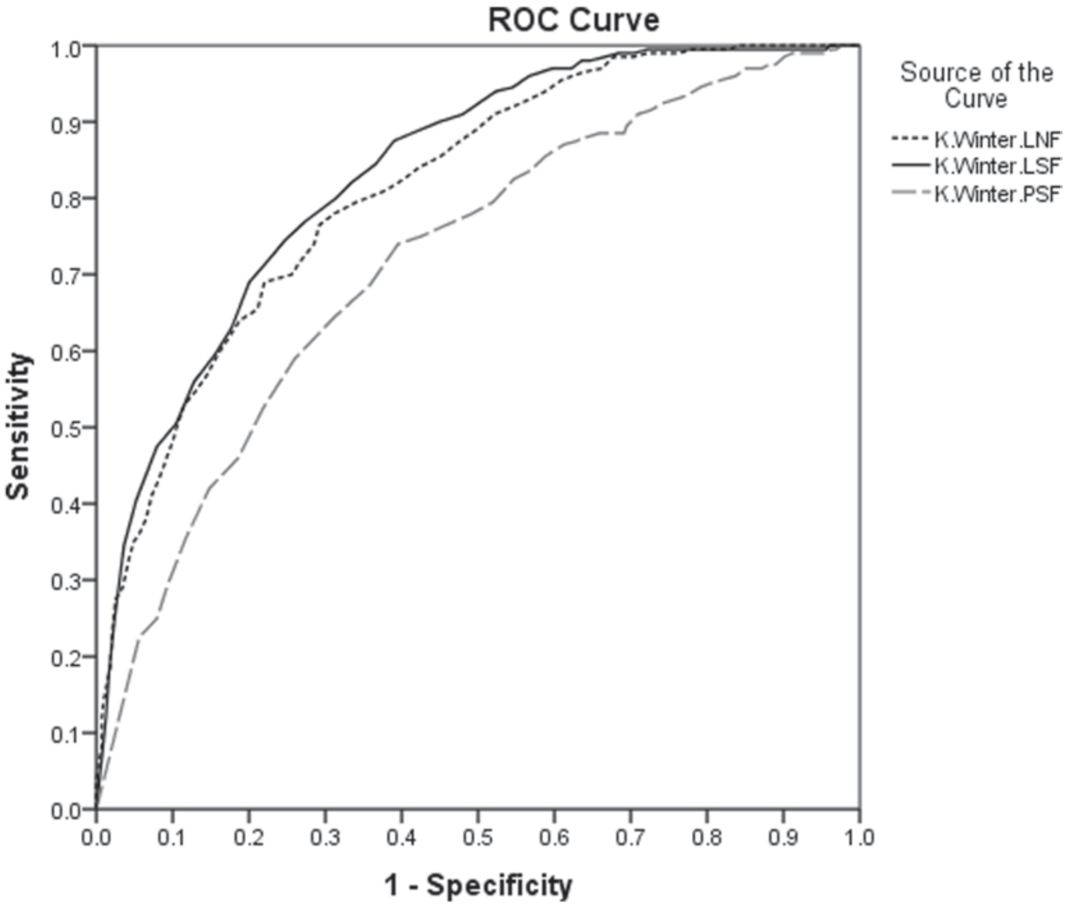


Fig. 2 ROC curves for kindergarten screening measures. ROC receiver operating characteristic

Table 1 Classification accuracy of single-predictor screening models

	Measure	AUC	Cut score	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
Derivation sample (N = 755)	LSF	0.833	20	0.90	0.55	0.42	0.94	180	305	250	20
	LNF	0.814	43	0.91	0.48	0.39	0.94	182	265	290	18
	PSF	0.719	27	0.90	0.31	0.32	0.89	179	169	385	21
Cross-validation sample (N = 298)	LSF		20	0.89	0.65	0.31	0.97	40	165	88	5

LSF letter–sound fluency, *LNF* letter-naming fluency, *PSF* phoneme segmentation fluency, AUC area under the curve, *TP* true positive, *TN* true negative, *FP* false positive, *FN* false negative, *PPV* positive predictive value, *NPV* negative predictive value

overall accuracy across all potential cut points) at 0.833, followed by letter-naming fluency (LNF; 0.814), and phoneme segmentation fluency (PSF; 0.719). Coordinates of the curve were inspected to determine a cut score on each measure that resulted in sensitivity of approximately 0.90, and the classification accuracy results using these cut scores are summarized in Table 1. With sensitivity held to a minimum of 0.90, the measures demonstrated specificity values between 0.31

and 0.55. High rates of false-positive errors were observed across all measures, ranging from 250 to 385 false-positive errors in our sample of 755 screened students, and only 46–64 % of students were correctly classified overall. For positive predictive value (PPV), values ranged from 0.32 for PSF to 0.42 for LSF, indicating that students were slightly *more* likely (58–68 %) to meet the first-grade criterion if they *failed* the screen. Based on the rates of false positives, using these measures on their own to identify students for intervention, without any follow-up or assessment to rule out false-positive errors, is empirically unsupported.

Cross Validation In the above example, cut scores and the accuracy of the screening tool were evaluated *within the same sample*. In practice, the critical next step is to cross-validate the accuracy of these cut scores on a new sample of students to determine if they demonstrate similar levels of accuracy. Using a subsequent cohort of students ($N=298$) from the same school district, the degree to which the LSF cut score generated in the previous analysis resulted in similar levels of accuracy in classifying students according to the first grade R-CBM score of 40 (the base rate in this sample was 15 %) is evaluated. Results are reported in the last row of Table 1. With this new cohort, the previously derived cut score of 20 on LSF in the winter of kindergarten demonstrated similar classification accuracy to LSF performance in the derivation sample, with nearly identical sensitivity (0.89 vs. 0.90), and slightly stronger specificity (0.65 vs. 0.55). Improved specificity was likely due to a lower base rate in this sample (15 %) compared to the derivation sample (26.5 %; the implications of base rate differences will be discussed later). However, high rates of false positives were still observed, and PPV was only 0.31 indicating a small portion of students identified as at risk on the screen later failed the first-grade criterion. On the other hand, the NPV was very high (0.97), indicating that a very portion of students who were considered "not at risk" on the screen met the first-grade criterion. Given similar sensitivity and specificity, PPV will predictably (mathematically) go down with a lower base rate. That is why PPV and NPV are not comparable across studies with different risk base rates.

Evaluating the Accuracy of Screens Unfortunately, throughout research on screening tools, similar patterns of unacceptable classification accuracy are documented for single screening assessments assessed at a single point in time (Catts et al. 2009; Clemens et al. 2011; Fuchs et al. 2007; Gofreda and DiPerna 2010; Jenkins et al. 2007; Johnson et al. 2009; Riedel 2007; VanDerHeyden 2011). Subsequently, problems with current methods of deriving cut scores and evaluating screening tools are discussed, after which some suggestions are offered for how these methods may be reconsidered and improved.

Current Problems and Challenges in Determining Screening Accuracy and Whom to Treat

Determining "whom to treat" is probably the most challenging and error-prone step of the screening process, made difficult due to the ambiguity often reflected by current screening tools and methods. When screening within educational contexts, there are several variables that can interfere with both the process of establishing cut points, as well as evaluating the accuracy of screening measures.

Time Delay Between Screening and "Outcome" Screening for academic difficulties is challenged by the fact that the authors are usually trying to identify problems that have not yet occurred (that is the purpose of screening). Unlike signal detection research (where ROC curve analyses originated), which evaluated individuals' accuracy in detecting signals that were actually present at the time, an academic skills screening and the subsequent outcome evaluation can be separated by many months or even years. This is problematic for evaluating screening accuracy due to what can take place during the intervening time. Students' language and cognitive skills are naturally maturing, teachers are tasked with teaching, curricula are often designed to individualize instruction for struggling students, formal interventions may be put into place to remediate skill deficits, and parents may intervene to help a struggling child. These factors can have the largest impact on rates of false-positive errors, because a student deemed "at risk" on a screening, but responds favorably to instruction and is successful on the outcome assessment, would be considered a false-positive error in a contingency table. Screening researchers should document the instruction and intervention efforts that occur between screening and outcome assessment, and attempt to control for the effect of intervening events on decision accuracy, which would represent a new direction in screening research.

Cut Scores Imply a Black and White Dichotomy The risk of using cut scores to communicate students' status on the screening measure is the unintended implication that all students within each group share the same level of risk. Splitting a sample into two groups can imply homogeneity within groups, when in actuality, scores at different points of the distribution may have different probabilities of success.

For example, consider R-CBM, a commonly used screening measure for reading, being used in a third grade to identify students at risk for failing a high-stakes test later in the year. A third-grade student in the fall of the school year may read anywhere between 0 and more than 140 words per minute, and let us assume the school used a cut score of 40 words read correctly per minute. Taken purely at face value, the cut score specifies that a student scoring at 38 wcpm is as much at risk as is a student scoring 19 wcpm, and a student scoring 9 wcpm. Conversely, a student scoring 42 wcpm would not be considered at risk, nor would students scoring 57 wcpm or 136 wcpm. However, the authors know that measures of oral reading have demonstrated moderate positive correlations with state high-stakes assessments of reading comprehension of around 0.65 (Reschly et al. 2009), indicating that higher R-CBM are associated with higher state test scores, and vice versa. It is easy to predict outcomes when scores fall at the extreme ends of the distribution; the student with a score of 9 wcpm is clearly at risk, and the student with a score of 136 wcpm is likely not. Problems arise when attempting to assign a risk decision to students in close proximity to the cut score. Students who scored 38 wcpm and 42 wcpm fall on opposite sides of the cut score, but the likelihood of a successful outcome for these two students is much more similar compared to the likelihood of success for students scoring 38 wcpm and 9 wcpm. However, the at-risk/not at-risk dichotomization implies two very different decisions, both with significant implications. When cut scores are set to maximize sensitivity, students falling below the cut score (but still in proximity to it) will be the main source of the false-positive errors. This is not clearly evident for the novice user using a cut score in a screening assessment, nor is it often acknowledged by researchers. The two preceding issues, the latency between screening and outcome assessment and problems with cut-score dichotomies, may be mitigated by estimating risk based on predicted probability and gated screening procedures which are discussed later in the chapter.

Difficulty in Extrapolating and Generalizing Screening Results to Other Settings Although preestablished cut scores are convenient and provide a user-friendly means for educators to evaluate risk, benchmark targets established with one population can result in poor classification accuracy with another, as has been observed in some cases when applying DIBELS recommended cut scores (Goffreda et al. 2009; Hintze et al. 2003; Nelson 2008; Roehrig et al. 2008) and is evident in the case example provided earlier in this chapter. Populations can differ markedly in terms of their demographics, different outcome measures may be used, and different decision rules signifying successful or unsuccessful outcomes.

A significant issue in attempting to generalize screening results and evaluating screening measures is the effect of base rates. The prevalence of the screening target (e.g., state test failure) within a given population will affect the "accuracy" of a screening tool. Studies have used anything from the 8th percentile to the 36th percentile as outcome criteria (see Jenkins et al. 2007). In a population in which failure is very high, it is much easier for a screening tool to predict failure, and the accuracy of the screening measure may be overestimated. Conversely, if failure rates are very low, it becomes more difficult to accurately predict failure.

Hence, the goal of screening (i.e., "what is the problem we are trying to detect?") has direct implications on base rates. In some cases, screening has been conducted to identify students who will later demonstrate or be identified with a reading disability (e.g., Gilbert et al. 2012), or has been defined by achievement below a certain percentile or standard score across a set of standardized measures (e.g., Compton et al. 2006, 2010). In other cases, screening has sought to identify students at risk for failing a statewide high-stakes assessment (e.g., Roehrig et al. 2008; Silbertglitt and Hintze 2005). Depending on the school or district, failing a state assessment may be a high- or low-probability event. Likewise, the problems of defining learning disabilities notwithstanding (Fletcher et al. 2006; VanDerHeyden in press), the base rates of learning disabilities will be higher in some populations compared to others. Similarly, using the 8th percentile as the criterion will result in a lower base rate than using the 36th percentile, thus a screening tool will be more likely to correctly identify students at risk for failure when the 36th percentile is used as opposed to the 8th.

Improving our Evaluation of Screening and Determining Whom to Treat

If common practices of single-point-in-time screenings, screening cut scores, and resulting classification accuracies are problematic for truly evaluating screening accuracy and determining which students we identify for intervention, how can we improve screening processes in multi-tiered intervention systems? Next, several innovations that may be considered for improving screening are described.

Logistic Regression Logistic regression is used to evaluate the validity of continuous or dichotomous independent variables (e.g., screening measures) in predicting a dichotomous-dependent variable (e.g., outcome or criterion level of performance). Relevant to screening, logistic regression provides a method for determining the probability of a subsequent outcome based on scores on a measure or set of measures, as well as a way to evaluate the accuracy of combinations of measures. Logistic regression has been used in this manner in academic screening research (Catts et al. 2001; Clemens et al. 2011; Compton et al. 2006; Johnson et al. 2010; Speece et al. 2010).

Using Predicted Probabilities Instead of Cut Scores Given the previously discussed problems posed by the latency between screening and outcome assessment and the forced dichotomy imposed by cut scores, it may be more informative to evaluate risk based on locally derived probability estimates of a particular outcome based on a *specific score* on the continuously scaled screening measure. This may better communicate the severity of risk, the likelihood of success, and the uncertainty of the decision for some students and the need to utilize additional information. Locally derived predicted probability estimates reflect the probability of a particular outcome given the latency between the screening and outcome assessment, the measured contextual factors present in that location, and the specific tool used for evaluating outcomes.

Probability information is used and interpreted in daily lives. Meteorological systems use atmospheric data to estimate the likelihood of rain on a given day. Levels of systolic and diastolic blood pressure are used to gauge the risk of heart attack. In educational screening, using probability information instead of forced dichotomies associated with single cut scores may allow educators to better evaluate risk by communicating the severity of risk for some students, the likelihood of success for others, and the uncertainty of prediction for students scoring in the middle of the distribution.

As an example of how predicted probabilities may better inform screening decision-making,

Table 2 Sample of students (from sample of 755) used in ROC and logistic regression analyses

				Multiple-measure model			
Student	K winter LSF score	ROC-derived risk categories	LSF-predicted probability	LSF score	LNF score	Predicted probability	*Met grade 1 criteria*?
1	68	"Not at risk"	0.999	68	70	0.999	Yes
2	47		0.996	47	41	0.987	Yes
3	34		0.977	34	66	0.985	Yes
4	27		0.941	27	18	0.838	Yes
5	22		0.890	22	61	0.948	Yes
6	20		0.860	20	51	0.910	Yes
7	19	"At risk"	0.843	19	64	0.940	Yes
8	19		0.843	19	28	0.789	No
9	15		0.757	15	60	0.902	Yes
10	15		0.757	15	34	0.768	No
11	13		0.704	13	56	0.868	Yes
12	13		0.704	13	22	0.633	No
13	11		0.644	11	52	0.825	Yes
14	11		0.644	11	14	0.512	Yes
15	5		0.443	5	41	0.640	Yes
16	5		0.443	5	1	0.268	No
17	3		0.377	3	37	0.559	Yes
18	3		0.377	3	2	0.241	No
19	0		0.287	0	50	0.618	Yes
20	0		0.287	0	0	0.183	No

$N=755$. LSF and LNF scores were collected in the winter of kindergarten. "Predicted probability" represents the probability of meeting the first-grade reading fluency criterion of 40 words per minute (i.e., 25th percentile)
ROC receiver operating characteristic, *LSF* letter–sound fluency, *LNF* letter-naming fluency

the authors return to our sample used earlier in the ROC curve analyses. A logistic regression model was run using LSF as the predictor of first-grade reading status, and the predicted probabilities of attaining the 40 words per minute criteria associated with each LSF score were saved. To illustrate, 20 students from the full sample of 755 were sampled and their scores are displayed in Table 2, rank-ordered from highest to lowest according to their LSF score in the winter of kindergarten. The column directly next to this score reflects the at-risk/not at-risk category in which each student was placed based on the dichotomous categorization using the 20 LSF cut score derived in the earlier ROC analysis. The next column to the right indicates the probability of passing the 40 wcpm criterion based on each LSF score. Of interest is the range of probabilities observed across the range of scores that fall within the "at-risk" range. For example, even a score of 15, earned by students 9 and 10, has a 76 % chance of meeting the first-grade criterion.

These data reveal why false-positive errors are so prevalent in assessments of classification accuracy among screening measures, as students in proximity to the cut score may still have a very high likelihood of actually meeting the outcome target. For example, students with LSF scores of 19 and 20 have a likelihood of 0.84 and 0.86, respectively, of attaining the first-grade criterion despite being on either side of the cut score. Instead of an all-or-nothing implication of above/below a cut score, probability values can provide a way to communicate the continuum of risk associated with scores across the distribution of a continuous variable, such as when the probability of success is highly likely or highly unlikely, and perhaps more importantly, indicate when more information is needed (e.g., probability between 0.40 and 0.70) to make a better decision regarding the allocation of intervention services. Tables can be created,

Table 3 Score ranges and associated probability of successfully meeting first-grade reading criterion from derivation sample, and percentages of students in cross validation cohort that met the first-grade criterion

LSF score	Predicted probability from derivation sample	% of students in subsequent cohort that met first-grade criterion	*N*
35+	≥0.98	99%	82
30–34	0.96–0.98	100%	25
25–29	0.92–0.96	95%	39
20–24	0.86–0.91	92%	24
15–19	0.76–0.84	82%	38
10–14	0.61–0.73	81%	32
5–9	0.44–0.58	68%	25
0–4	0.29–0.41	42%	33

N of derivation sample = 755; *N* of cross-validation cohort = 298
LSF letter–sound fluency

such as Table 3, that summarize the score ranges on the screening measure and the probability of a given outcome observed in prior cohorts in that school or district. When using a table of probability values like those displayed in Table 3, categorical decisions (e.g., identify for intervention or further testing) will still result. But the importance of probability values rests in their ability to better communicate to educators the level of risk they are dealing with for each student.

Predicted probabilities may also provide a better sense of risk status based on the characteristics of the local context, such as the population, the curriculum, interventions in place, and the tiered model of instruction. It should be noted that this process does not negate the previously discussed concern regarding the intervening factors occurring between the screen and the criterion assessment that can have an effect of "creating" false positives. However, by using probabilities empirically derived from the sample in which those same factors are present (i.e., same teachers, curricula, intervention system), decisions now might be based on a *continuum* of risk. This may enable clearer determinations of the students that are clearly at risk (e.g., probability < 0.40), more clearly not at risk (e.g., probability ≥ 0.85), and which students are too difficult to classify based on the results of a static screen and for whom more information may be needed (e.g., probability in proximity to chance levels, such as 0.40–0.70). This information might be gathered in second stage of screening, and may improve decisions regarding the delivery of costly intervention services. Similar thinking has been applied in medicine in which the probability of disease is considered in light of test results, such as Pauker and Kassirer's (1980) threshold approach to determining when to begin treatment, when treatment is not necessary, and when uncertainty dictates the need to conduct additional testing (VanDerHeyden in press).

Probability values can be evaluated with subsequent cohorts to determine their consistency in predicting outcomes in a following year. Table 3 also includes the percentage of students from our cross-validation cohort that met the first-grade criterion when obtaining the associated LSF scores in kindergarten. The logistic regression can be updated each year to ensure that probability estimates are based on contextual characteristics currently observed in the setting (e.g., instructional curricula, interventions in place, latency between screen and criteria).

Multiple Measures Evidence from RTI implementation studies indicates variability in the number of screening measures schools use, with some reporting using three or more measures (Mellard et al. 2009), and others reporting reliance on single measures exclusively (Jenkins et al. 2013). The use of multivariate screening approaches has been recommended (Jenkins et al. 2007), and studies indicate that improved accuracy is obtained when measures are combined (Clemens et al. 2011; Compton et al. 2006; Fuchs et al. 2007; Johnson et al. 2010; O'Connor and Jenkins 1999; Speece et al., 2011; Shapiro et al. 2008; Speece et al. 2010).

The use of multiple-measure screening batteries must be considered carefully however, and it is not wise to indiscriminately recommend that multiple-measure approaches should be used in every case. Although some additional measures may be very brief and inexpensive (e.g., brief teacher ratings), additional individually administered multiple measures waste resources and instructional time. Additionally, using multiple measures for screening can be confusing, as it can be difficult to make sense of several different scores (which can occur when schools administer multi-measure batteries such as DIBELS). Any consideration of using multiple measures should first determine if adding measures represents added value in accuracy. Logistic regression provides the ability to combine multiple predictors (multiple screening tools) and determine which are statistically significant in the prediction of the target outcome. Those that are not significant can be viewed as having a negligible effect, at least statistically, on improving decision-making accuracy.

An example of using logistic regression to evaluate multiple-measure models is provided. Table 4 displays the results for each model when kindergarten LSF, LNF, and PSF were used alone (models 1–3), as well as models that combined predictors (models 4–6)[1].

When using LSF, LNF, and PSF separately as single predictors (models 1–3), each variable was significant in the logistic regression. Consistent with the earlier ROC curve analyses, LSF was the strongest predictor of first-grade reading fluency status based on regression weight, odds ratio, and model fit. Given that LSF was the strongest single predictor, the authors next tested whether adding additional measures resulted in improvements in predicting first-grade reading status over LSF alone (model 1). LNF was added in model 4, which resulted in significant improvements in fit over model 1 (diff = 19.17, $p < 0.001$), and both LSF and LNF were statistically significant predictors in the model. Combining LSF and PSF (model 5) did not result in improved fit over model 1 (diff = 3.73, $p = 0.053$), and PSF was not statistically significant as a predictor. Model 6 combined all three predictors (LSF, LNF, and PSF); this model was an improvement over model 1 (diff = 22.43 (2), $p < 0.001$), but did not significantly improve upon model 4 (diff = 3.26, $p = 0.071$). LSF and LNF were significant in this model, but PSF was not. Therefore, the best-fitting, most parsimonious model included two predictors, LSF and LNF (model 4).

Evaluating the statistical significance of predictors in a logistic regression model is a good first step, but next it is important to determine if the combination of measures improves accuracy appreciably over more parsimonious models. Using the logistic regression model 4 that combined LSF and LNF as predictors, the predicted probabilities of passing the first-grade criterion for every observed combination of scores on LSF and LNF were saved. An ROC curve for this model using the predicted probabilities as the predictor variable was then generated. This model had an AUC of 0.842 (slightly stronger than LSF alone), and setting sensitivity at 0.90, it demonstrated specificity of 0.58. This slight increase in specificity prevented 19 false-positive errors. This is a minor improvement in accuracy over using LSF alone, to be sure. Decisions must be made whether adding a measure to the screening battery is a better use of resources than placing a student who is not at risk in intervention. As Fuchs et al. (2011b) demonstrated, the costs of intervention may justify additional screening. As opposed to lengthening the universal screening battery, multiple or additional measures may be best allocated for obtaining additional information for students who may be at risk, or for instructional placement or planning purposes. This additional assessment need only take place with

[1] The p value indicates the statistical significance of that predictor in the model for predicting the subsequent first-grade reading outcome (40 words per minute), and the odds ratio indicates the percentage change in likelihood of the subsequent outcome given every point increase in the predictor (e.g., the odds ratio of 1.15 for LSF means that every 1 point increase in LSF equals an increase of 15 % probability of meeting the first grade criterion). The log likelihood value (expressed as −2LL; log likelihood is multiplied by −2 to derive a value equivalent to a chi square distribution) can be used to evaluate model fit and compare models. Smaller −2LL values indicate better-fitting models.

Table 4 Logistic regression results comparing single- and multiple-predictor models in predicting first-grade reading status

Model	Predictor(s)	β	SE	Wald	p	Odds ratio (95 % CI)	$-2LL$
1	Constant	−0.91	0.17	30.59	<0.001		
	LSF	0.14	0.01	124.76	<0.001	1.15 (1.12–1.17)	649.97
2	Constant	−1.53	0.22	50.62	<0.001		
	LNF	0.08	0.01	139.80	<0.001	1.09 (1.07–1.10)	671.23
3	Constant	0.17	0.13	1.92	0.17		
	PSF	0.07	0.01	57.08	<0.001	1.07 (1.05–1.09)	793.83
4	Constant	−1.494	0.22	46.43	<0.001		
	LSF+	0.09	0.02	34.13	<0.001	1.09 (1.06–1.13)	
	LNF	0.04	0.01	18.25	<0.001	1.04 (1.02–1.06)	630.80
5	Constant	−1.00	0.17	33.18	<0.001		
	LSF+	0.13	0.01	96.87	<0.001	1.14 (1.11–1.17)	
	PSF	0.02	0.01	2.86	0.09	1.02 (1.00–1.04)	646.24
6	Constant	−1.57	0.23	48.49	<0.001		
	LSF+	0.08	0.02	26.19	<0.001	1.09 (1.05–1.12)	
	LNF+	0.04	0.01	17.85	<0.001	1.04 (1.02–1.06)	
	PSF	0.02	0.01	2.44	0.11	1.02 (1.00–1.04)	627.54
Model	Predictor(s)	β	*SE*	Wald	p	*Odds Ratio (95 % CI)*	$-2LL$
1	Constant	−0.91	0.17	30.59	<0.001		
	LSF	0.14	0.01	124.76	<0.001	1.15 (1.12–1.17)	649.97
2	Constant	−1.53	0.22	50.62	<0.001		
	LNF	0.08	0.01	139.80	<0.001	1.09 (1.07–1.10)	671.23

$N = 755$.

LSF letter–sound fluency, *LNF* letter-naming fluency, *PSF* phoneme segmentation fluency, $-2LL$=log likelihood (multiplied by −2 in order to compare to chi-square distribution), *CI* confidence interval. Model 1 versus model 4: diff=19.17, $p<0.001$; model 1 versus model 5: diff=3.73, $p=0.053$; model 1 versus model 6: diff=22.43 with 2 df, $p<0.001$; model 4 versus model 6: diff=3.26, $p=0.071$; model 5 versus model 6: diff=18.70, $p<0.001$).

Table 5 Example multiple-measure matrix summarizing winter kindergarten LSF and LNF scores and associated probabilities of meeting the subsequent first-grade criterion

LSF score	LNF score 55+	40–54	35–39	30–34	25–29	20–24	15–19	10–14	5–9	0–4
25+	≥0.95	0.92–0.95	0.90–0.93	0.87–0.91	0.87–0.89					
20–24	0.93–0.95	0.87–0.93	0.84–0.89	0.82–0.88	0.79–0.86					
15–19	0.90–0.95	0.81–0.91	0.78–0.85	0.75–0.83	0.70–0.79	0.67–0.79	0.65–0.70			
10–14	0.87–0.90	0.74–0.84	0.69–0.78	0.67–0.74	0.60–0.71	0.55–0.63	0.50–0.61	0.47–0.58		
5–9		0.64–0.74	0.59–0.69	0.54–0.67	0.49–0.61	0.44–0.55	0.39–0.50	0.34–0.45	0.31–0.37	
0–4		0.55–0.66	0.56–0.57	0.49–0.53	0.42–0.49	0.34–0.42	0.29–0.41	0.26–0.36	0.22–0.30	0.18–0.26

Empty cells = too few students achieved that combination of scores to generate a probability value
LSF letter–sound fluency, *LNF* letter-naming fluency

a subset of students; resources need not be spent on students who are clearly not at risk (or who are clearly at risk).

It must be understood, however, that to this point in the evaluation of the LSF + LNF model the authors have relied on the 2 × 2 contingency for evaluating its accuracy. Predicted probabilities may again better inform decisions made based on the screening data with multiple measures. To illustrate, Table 2 includes the predicted probability results of model 4 (LSF + LNF). The LSF and LNF scores for the same 20 students are provided, as well as the probability of each student successfully meeting the first-grade criterion based on each *combination* of LSF and LNF scores. In the far right column, it is also indicated whether each student sampled actually met the first-grade criterion.

Probabilities change with the inclusion of additional predictors. For example, consider students #15 and 16. Both students had scores of 5 on LSF, and when LSF is considered alone, a score of 5 is associated with a 44 % chance of successful outcomes on our first-grade criterion. But consider their LNF scores simultaneously; student 15 (LNF = 41) now has a probability of 64 % for successful outcomes, compared to 27 % for student 16 (LNF = 1). Incidentally, student 15 met the first-grade criterion, and student 16 did not. Other interesting patterns can be observed across the sample, and illustrate the potential for a stronger estimate of risk status when using the probabilities associated with a combination of scores. A matrix of combined scores might then be created, as displayed in Table 5, and educators might use these empirically derived probabilities to compare students' scores from subsequent screening assessments.

Stage 2 Screening: Dynamic Assessment Although it has many conceptualizations and definitions, dynamic assessment (DA) generally refers to an assessment activity that evaluates learning potential using instructional tasks, prompts, or performance feedback, and is implemented for the purpose of determining learner characteristics that may not be as readily gained from conventional assessment (Wagner and Compton, 2011). In addition to assessing a student's level of functioning at the time of the assessment, rate of growth within the assessment is of particular importance in DA (Fuchs et al. 2011a).

Typically, DA includes presenting content unfamiliar to the student, using a set of graduated prompts or scaffolds, during which the examiner evaluates the rate of acquisition and level of support needed to reach a pre-determined mastery criterion. Several recent publications have highlighted the potential for DA to improve the screening process by providing a means to potentially reduce false-positive decision errors. For example, with beginning first-grade students, Fuchs et al. (2011a) found that a DA of decoding skills explained unique variance in the prediction of subsequent reading difficulty over static assessments, suggesting that it may be a viable option in determining reading risk over having to wait several weeks to evaluate progress during an intervention. Additional studies have found support for DA to improve screening decisions across academic areas, particularly with regard to reducing false-positive errors (Elleman et al. 2011; Jeltova et al. 2011; Fuchs et al. 2011b; Seethaler et al. 2012; Sittner-Bridges and Catts 2011; Swanson and Howard 2005). DA is time intensive, and it is not being advocated as a method for universal screening. Rather, DA might be considered as a step carried out with a subset of students considered to be at risk and for whom more information is required. It also should be acknowledged that DA may offer an improvement to static assessment in some cases but not others (e.g., preschool phonological awareness; see Coventry et al. 2011; Kantor et al. 2011).

Stage 2 Screening: Progress Monitoring The concept of using progress monitoring as a method for screening is not a new one (Marston et al. 1984). More recently, however, studies have evaluated the implementation of a short-term period of progress monitoring as part of a multi-stage screening process. Following the universal screen, a subset of at-risk students would be monitored on a frequent basis (i.e., weekly, biweekly) across several weeks while intervention is implemented. Rate of growth is then evaluated to identify students unresponsive to instruction and in need of additional, more intensive interventions. Compton et al. (2006, 2010) found that, when added to the results on an initial static screening battery, a 6-week period of progress monitoring using word identification fluency resulted in improved accuracy in identifying subsequent reading disability status with first-grade students. Similar results were obtained by Gilbert et al. (2012). Other work in reading found that growth estimates were indeed significant and unique predictors of later reading status in addition to static predictors; however, the modest improvements in accuracy gained by the growth estimates did not justify their implementation or delay of intervention when multivariate models of static screens performed similarly (Speece et al. 2010, 2011).

Data Mining and Classification Tree Analyses It may be instructive to consider methods used in other fields that are concerned with the classification of individuals in anticipation of a subsequent outcome. For example, Chi-squared Automatic Interaction Detection (CHAID) is an algorithmic classification tree method of data mining that can be used for prediction and classification. Similar to regression analyses, CHAID can be used to model interactions between predictors and optimally predict a dependent dichotomous variable, and is used in banking and insurance to determine the characteristics of loan applicants that are/ are not likely to default on a loan. Applied to educational screening, data mining techniques might identify sets of indicator skills at school entry as well as in subsequent grades that are associated with greater risk for poor outcomes. Compton et al. (2006) utilized classification tree analyses to determine a set of variables effective for reading disability screening. Data mining and classification tree analyses such as CHAID require large data sets, which are becoming more available via state-wide data collection efforts and national longitudinal data sets.

Summary, Implications, Next Steps, and Recommendations for Practice

Work to date has provided a good foundation to understanding potentially effective tools and methods for academic skills screening, but there is much room for improvement. Next, a set of suggestions for advancing our knowledge and understanding of effective academic screening practices is offered.

Improve Screening Content CBM tools have characteristics that make them attractive for screening, but work should continually seek to refine and improve assessments, as well as consider previously untapped skills or assessment technologies. Much may be learned from the research on language, vocabulary, attention, and behavior regulation in preschool and early elementary school; skills that have significant and broad reaching effects later academic achievement and life-course outcomes in general.

Explore New Methods of Data Collection Beyond 1:1 Direct Assessments Group-administered assessment holds significant logistical advantages for universal screening, and may improve overall accuracy when combined with a brief direct assessment. Computer-administered and computer-adaptive approaches offer improved efficiency, and may provide more detailed diagnostic information that may improve classification accuracy, and provide more information for instructional planning. Researchers must evaluate their accuracy relative to individually administered measures. Screening assessment should also not neglect the input of teachers, who may be able to provide information on student achievement and skills in ways that individually administered measures may not.

Thinking Differently About Screening "Accuracy" Risk Determination

Within educational contexts, false-positive errors are inevitable. In fact, in any setting where the screening tool and criterion outcome assessment are separated in time, teachers are tasked with teaching, students are learning and maturing, intervention efforts are put into place, and parents seek extra help, false-positive errors should be expected. Moreover, as RTI systems become more prevalent and core instruction and intervention systems improve, one should expect the rates of false-positive errors within our schools as identified on screening measures to increase. In an educational environment, a goal to completely or nearly eliminate false-positive errors is unrealistic. Furthermore, an at-risk student with a strong RTI, what researchers may deem a "false positive" and a failure of a screen, may be considered by teachers to be an educational victory. Perhaps researchers should adjust their way of thinking about false-positive errors, and "screening accuracy" in educational contexts, particularly RTI.

Consider Screening as a Process The preceding discussion speaks to the need for appropriate conceptualizations of universal screening and the screening process in general. All of the burden to "get it right" should not be placed solely on the shoulders of universal screening. Consistent with the models described by Fuchs et al. (2011b) and Gilbert et al. (2012), which incorporate multistage screening and/or progress monitoring data embedded within the screening system, screening should be thought of as an ongoing process. The authors advocate that the aim of the universal screen should be to quickly and inexpensively identify the subset of students who are at low risk for later failure, and thus excused from further testing. Valid group-administered measures with good sensitivity that are efficient to administer with large groups of students are ideal at this stage. Computer-adaptive assessments might be considered here given that they can be administered in group settings, the data are readily available and automatically scored, and they may be ideal for identifying the students that are not at risk. Teacher ratings might also be incorporated with academic assessments to help improve identification.

Using predicted probabilities generated from logistic regression with prior cohorts of students collected locally may help better inform decisions made regarding screening, by identifying the students for whom more information is needed to make a decision. With this subset of students, additional assessments may be considered to (a) better ensure that students are truly demonstrating a pattern of skills that indicates they are indeed at risk if nothing were changed and (b) identify skill areas in which intervention may be needed. Measures shown to improve classification accuracy (through analyses such as logistic regression) and that improve diagnostic decision-making should be considered here, much like Vaughn and Fletcher's (2012) use of R-CBM, follow state test results. This stage may include additional assessment methods, such as DA, or diagnostic measures using assessments more targeted and specific to the area of difficulty. Some intervention curricula include placement assessments to determine the appropriateness of the curriculum for the students' needs, and the entry point in the curriculum which may be most appropriate, and these placement assessments may be excellent options at this stage.

For students placed in supplemental interventions, frequent progress monitoring should be conducted to determine responsiveness to instruction and progress toward outcomes. In addition to informing instruction, progress monitoring also serves the purpose of further identifying false-positive errors. False-positive errors represent a costly use of resources (Fuchs et al. 2011b), but with frequent progress monitoring false-positive errors need not remain in intervention for long. Students demonstrating rapid growth, or early achievement of middle- or end-of-year goals, should be considered for reducing the intensity of the intervention or discontinuing intervention altogether, if appropriate.

The Need for Continued Technical Innovation, Training, and Support Combining measures may result in improved accuracy over single screening tools, but analysis of such screening models requires technology and training. To better prepare school personnel, training programs for school psychologists might integrate instruction on classification accuracy and logistic regression into their basic courses on statistics. The authors also advocate that online data systems (e.g., DIBELS, AIMSweb, EasyCBM) incorporate methodologies such as logistic regression and probability analyses into their software to allow educators to better interpret screening assessment with statistics that are better suited to educational contexts. Some systems exist that are making the effort to incorporate better data analytics to inform screening decisions, such as the TIES system in Minnesota which uses logistic regression to estimate probabilities of passing state-wide assessments (Benjamin Silberglitt, personal communication, July 1, 2013).

Consider the Context Screening in later elementary, middle, and high school can (and should) look different from screening in early elementary school. Screening may be less frequent for older students, and extant assessment data, such as state test results, may be sufficient for universal screening purposes. For example, Denton et al. (2011) found the best predictors of high-stakes test outcomes were students' scores on the assessment the prior year, and additional assessments (e.g., R-CBM, Maze, multiple-choice comprehension) did little to improve the prediction. For the purposes of identifying students for tiered intervention, Vaughn and Fletcher (2012) found that using state test scores and adding a 1-min measure of oral reading fluency helped to accurately divide students into subgroups based on instructional needs (e.g., fluency, decoding, comprehension).

Consider the Costs Schools must balance the need for assessment against its costs. Costs of screening in terms of resources and instructional time must be carefully weighed against the costs of intervention per student. Additionally, thinking that more assessment is better and collecting too much data can cause "paralysis by analysis." The processes and procedures that have been advocated for place a premium on efficiency (e.g., using 1–2 measures for universal screening), eliminating students clearly not at risk from further testing, and saving additional assessment steps for smaller subsets of students. Probability indices may provide educators with a more detailed picture of risk status, informing more timely and targeted decisions and allowing them to get on with teaching.

References

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment system*. Eugene: University of Oregon, Behavioral Research and Teaching.

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology, 47*(4), 905–915.

Anderson, D., Alonzo, J., & Tindal, G. (2010). *easyCBM Mathematics Criterion Related Validity Evidence: Washington State Test* (Technical Report No. 1010). Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Jasmine Park, B., Irvin, P. S., Alonzo, J., & Tindal, G. (2011). *Diagnostic efficiency of EasyCBM Reading: Washington state* (Technical Report #1107). Eugene: Behavioral Research and Teaching, University of Oregon.

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Restetar, J. L., Slider, N. J., & Williams, K. L. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33,* 218–233.

Baglici, S. P., Codding, R. S., & Tryon, G. (2010). Extending the research on tests of early numeracy: Longitudinal analyses over two years. *Assessment for Effective Intervention, 35,* 89–102.

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., Reschly, D., Schrag, J., & Tilly, D. (2006). *Response to intervention: Policy considerations and implementation*. Alexandria: National Association of State Directors of Special Education.

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23*(1), 43–55.

Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review, 40*(1), 23–38.

Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools, 40*(2), 161–173.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*(1), 38–50.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163–176.

Catts, H. W., Compton, D., Tomblin, J. B., & Sittner Bridges, M. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*(1), 166.

Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of recent research: Curriculum based measurement of math computation. *Assessment for Effective Intervention, 33,* 198–205.

Clarke, B., Nese, J. F., Alonzo, J., Smith, J. L. M., Tindal, G., Kame'enui, E. J., & Baker, S. K. (2011). Classification accuracy of easyCBM first-grade mathematics measures: Findings and implications for the field. *Assessment for Effective Intervention, 36,* 243–255.

Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. *School Psychology Quarterly, 26,* 231–244.

Clemens, N. H., Hagan-Burke, S., Luo, W., Cerda, C. A., Blakely, A., Frosch, J., Gamez, B., & Jones, M. (in press). Investigating the validity of a computer-adaptive assessment of early reading in kindergarten. *School Psychology Review*.

Clemens, N. H., Oslund, E. L., Simmons, L. E., & Simmons, D. (2014). Assessing spelling in kindergarten: Further comparison of scoring metrics and their relation to reading skills. *Journal of School Psychology, 52,* 49–61.

Cocker, D. L., & Ritchey, K. D. (2010). Curriculum based measurement of writing in kindergarten and first grade: An investigation of production and qualitative scores. *Exceptional Children, 76*(10), 175–193.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98,* 394–409.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102,* 327–340.

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45*(1), 79–95.

Coventry, W. L., Byrne, B., Olson, R. K., Corley, R., & Samuelsson, S. (2011). Dynamic and static assessment of phonological awareness in preschool: A behavior-genetic study. *Journal of Learning Disabilities, 44*(4), 322–329. doi:10.1177/0022219411407862.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R. Jr., Ruby, M., Crevecoeur, Y. C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness, 3*(2), 93–120.

Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: Comparison of actual and predicted performances. *School Psychology Quarterly, 13*(1), 8–24.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184–192. doi:10.1177/00224669030370030801.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48*(4), 368–371.

DiPerna, J. C., & Elliott, S. N. (1999). Development and validation of the academic competence evaluation scales. *Journal of Psychoeducational Assessment, 17*(3), 207–225.

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., ... & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading, 15*(2), 109–135.

Eaves, R. C., Williams, P., Winchester, K., & Darch, C. (1994). Using teacher judgment and IQ to estimate reading and mathematics achievement in a remedial-reading program. *Psychology in the Schools, 31*(4), 261–272. doi:10.1002/1520-6807.

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools, 43*(3), 247–265. doi:10.1002/pits.20147.

Eckert, T. L., Koenig, E. A., Hier, B. O., & Arbolino, L. A. (2013). The role of teacher perspectives in diagnostic and program evaluation of academic skills. In R. Brown-Chidsey & K. J. Andren (Eds.), *Assessment for intervention: A problem-solving approach* (2nd ed., pp. 62–76). New York: Guilford Press.

Egan, J. P., Clarke, F. R., & Carterette, E. C. (1956). On the transmission and confirmation of messages in noise. *The Journal of the Acoustical Society of America, 28,* 536.

Elleman, A. M., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2011). Exploring dynamic assessment as a means of identifying children at risk of developing comprehension difficulties. *Journal of Learning Disabilities, 44*(4), 348–357.

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *The Journal of Special Education, 34*(3), 140–153. doi:10.1177/002246690003400303.

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of school psychology, 49*(2), 197–216.

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102*(6), 453–462.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. (2006). *Learning disabilities*. New York: Guilford Press.

Formative Assessment System for Teachers. (2013). *Formative Assessment system for teachers*. University of Minnesota, Author.

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing, 18*(2), 129–155.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*(1), 45–58.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195–203.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20–28.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*(4), 216–228.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness-to-intervention. *Exceptional Children, 72,* 311–330.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011a). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*(4), 339–347.

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (2011b). Two-stage screening for math problem-solving difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities, 44,* 372–380.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78,* 263–279.

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based writing and rating-based measures of written expression for elementary school students. *School Psychology Review, 35,* 435–450.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293–304.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009a). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention in the primary grades. A practice guide. Washington,* DC: Institute of Education Sciences.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009b). *Assisting students struggling with mathematics: Response to intervention (rti) for elementary and middle schools. NCEE, 4060*. Washington, DC: Institute of Educational Services, U.S. Department of Education.

Gilbert, J. K., Compton, D. L., Fuchs, D., & Fuchs, L. S. (2012). Early screening for risk of reading disabilities: Recommendations for a four-step screening system. *Assessment for Effective Intervention, 38*(1), 6–14.

Glover, T., & Albers, C. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135.

Goffreda, C. T., & DiPerna, C. (2010). An empirical review of psychometric evidence for the dynamic indicators of basic early literacy skills. *School Psychology Review, 39*(3), 463–483.

Goffreda, C. T., DiPerna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the dynamic indicators of basic early literacy skills (DIBELS). *Psychology in the Schools, 46*(6), 539–552.

Gonzalez, J. E., Goetz, E. T., Hall, R. J., Payne, T., Taylor, A. B., Kim, M., & McCormick, A. S. (2011). An evaluation of Early Reading First (ERF) preschool enrichment on language and literacy skills. *Reading and Writing, 24*(3), 253–284.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene: Institute for the Development of Education Achievement. http://dibels.uoregon.edu.

Goodenough, D. J. (1975). The use of ROC curves in testing the proficiency of individuals in classifying pneumoconiosis. *Radiology, 114*(2), 472–473.

Graham, S., & Perin, D. (2007). What we know, what we still need to know: Teaching adolescents to write. *Scientific Studies of Reading, 11,* 313–335.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). Meta-analysis of writing instruction for students in elementary school grades. *Journal of Educational Psychology, 104*(4), 879–896.

Graney, S. B. (2008). General education teacher judgments of their low-performing students' short-term reading progress. *Psychology in the Schools, 45*(6), 537–549.

Graney, S. B., Martínez, R. S., Missall, K. N., & Aricak, O. T. (2010). Universal screening of reading in late elementary school: R-CBM versus CBM Maze. *Remedial and Special Education, 31,* 368–377.

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *Test of silent contextual reading fluency*. Austin: PRO-ED.

Hagan-Burke, S., Kwok, O. M., Zou, Y., Johnson, C., Simmons, D., & Coyne, M. D. (2011). An examination of problem behaviors and reading outcomes in kindergarten students. *The Journal of Special Education, 45,* 131–148.

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32,* 228–240.

Harber, J. R. (1981). Learning disability research: How far have we progressed? *Learning Disability Quarterly, 4*(4), 372–381.

Hasselbring, T. S. (1984). Computer-based assessment of special-needs students. *Special Services in the Schools, 1*(1), 7–19.

Hecht, S. A., & Greenfield, D. B. (2001). Comparing the predictive validity of first grade teacher ratings and reading-related tests on third grade levels of reading skills in young children exposed to poverty. *School Psychology Review, 30*(1), 50–69.

Hintze, J., & Silbertglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high stakes testing. *School Psychology Review, 34,* 372–386.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *School Psychology Review, 32*(4), 541–556.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*(1), 127–155.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297–313.

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2011). Making instruction and assessment responsive to diverse students' progress: Group-administered dynamic assessment in teaching mathematics. *Journal of Learning Disabilities, 44*(4), 381–395.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*(4), 582–600.

Jenkins, J. R., Schiller, E., Blackorby, J., Thayer, S. K., & Tilly, W. D. (2013). Responsiveness to intervention in reading: Architecture and practices. *Learning Disability Quarterly, 36*(1), 36–46.

Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review, 34,* 27–44.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24,* 174–185.

Johnson, E. S., Jenkins, J. R., & Petscher, Y. (2010). Improving the accuracy of a direct route screening process. *Assessment for Effective Intervention, 35,* 131–140.

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of

mathematics proficiency in third grade. *School Psychology Review, 39,* 181–195.
Kantor, P. T., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2011). Comparing two forms of dynamic assessment and traditional assessment of preschool phonological awareness. *Journal of Learning Disabilities, 44*(4), 313–321.
Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300.
Kegel, C. A., van der Kooy-Hofland, V. A., & Bus, A. G. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences, 19*(4), 549–554.
Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review, 37*(3), 374–390.
Kenny, D. T., & Chekaluk, E. (1993). Early reading performance: A comparison of teacher-based and test-based assessments. *Journal of Learning Disabilities, 26*(4), 227–236.
Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of teacher judgments of preschoolers' math skills. *Journal of Psychoeducational Assessment, 30*(2), 148–159.
Landry, S., Assel, M. A., Gunnewig, S. B., & Swank, P. R. (2007). *MClass: CIRCLE.* New York: Wireless Generation.
Lane, K. L., Little, M. A., Casey, A. M., Lambert, W., Wehby, J., Weisenbach, J. L., & Phillips, A. (2009). A comparison of systematic screening tools for emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 17*(2), 93–105. doi:10.1177/1063426608326203.
Lee, J. M., Clark, W. W., & Lee, D. M. (1934). Measuring reading readiness. *The Elementary School Journal, 34*(9), 656–666.
Lee, Y. S., Lembke, E., Moore, D., Ginsburg, H. P., & Pappas, S. (2012). Item-level and construct evaluation of early numeracy curriculum-based measures. *Assessment for Effective Intervention, 37*(2), 107–117. doi:10.1177/1534508411431255.
Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research and Practice, 24*(1), 12–20.
Lembke, E. S., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention, 28*(3&4), 23–35.
Lembke, E. S., Hampton, D., & Beyers, S. H. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools, 49*(3), 257–272.
Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*(5), 451–459.
Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools, 40,* 379–390.
Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology, 47,* 315–335.
Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education, 18,* 109–117.
Masterson, J. J., & Apel, K. (2010). Linking characteristics discovered in spelling assessment to intervention goals and methods. *Learning Disability Quarterly, 33,* 185–198.
Martin, S. D., & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools, 48*(4), 343–356.
Martinez, R. S., Missal, K. N., Bamonto-Graney, S., Aricak, T. O., & Clarke, B. (2009). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention, 34*(2), 116–125.
Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). TOSWRF: Test of silent word reading fluency: Examiner's manual. Austin: Pro-Ed.
McBride, J. R., Ysseldyke, J., Milone, M., & Stickney, E. (2010). Technical adequacy and cost benefit of four measures of early literacy. *Canadian Journal of School Psychology, 25,* 189–204.
McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21,* 471–490.
McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education, 41*(2), 68–84.
McMaster, K. L., & Campbell, H. (2008). Technical features of new and existing measures of written expression: An examination within and across grade levels. *School Psychology Review, 27*(4), 550–566.
McMaster, K. L., Du, X., & Petursdottir, A. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities, 42,* 41–60.
McMaster, K. L., Parker, D., & Jung, P. G. (2012). Using Curriculum-Based measurement for beginning writers within a response to intervention framework. *Reading Psychology, 33,* 190–216.
Meisinger, E. B., Bradley, B. A., Schwanenflugel, P. J., Kuhn, M. R., & Morris, R. D. (2009). Myth and reality of the word caller: The relation between teacher nominations and prevalence among elementary school children. *School Psychology Quarterly, 24*(3), 147–150.
Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-

monitoring practices in 41 local schools. *Learning Disabilities Research & Practice, 24*(4), 186–195.

Mercer, S., Martinez, R. S., Faust, D., & Mitchell, R. R. (2012a). Criterion-related validity of writing curriculum-based measurement (WCBM) in high school students. *School Psychology Quarterly, 27*(2), 85–95.

Mercer, S. H., Dufrene, B. A., Zoder-Martell, K., Harpole, L. L., Mitchell, R. R., & Blaze, J. T. (2012b). Generalizability theory analysis of cbm maze reliability in third-through fifth-grade students. *Assessment for Effective Intervention, 37*(3), 183–190.

Metz, C. E., Goodenough, D. J., & Rossmann, K. (1973). Evaluation of receiver operating characteristic curve data in terms of information theory, with applications in radiography. *Radiology, 109*(2), 297–303.

Missal, K., Mercer, S., Martinez, R. S., & Casebeer, D. (2012). Concurrent and predictive patterns and trends in performance on early numeracy curriculum-based measures in kindergarten and first grade. *Assessment for Effective Intervention, 37*(2), 95–106.

NCS Pearson. (2012). *Test of early literacy administration and scoring guide*. Bloomington: NCS Pearson.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the national early literacy panel*. Washington, DC: National Institute for Literacy.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Bethesda: National Institute of Child Health and Human Development.

Nelson, J. M. (2008). Beyond correlational analysis of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A classification validity study. *School Psychology Quarterly, 23*(4), 542–552.

Northwest Evaluation Association. (2013). Measures of Academic Progress. www.nwea.org.

Norwalk, K. E., DiPerna, J. C., Lei, P., & Wu, Q. (2012). Examining early literacy skill differences among children in head start via latent profile analysis. *School Psychology Quarterly, 27*(3), 170–183.

O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159–197.

Parker, C., & Ditkowsky, B. (2006). *Examining the predictive validity of the dynamic indicators of vocabulary skills*. Technical Report, 4. Westport.

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *The New England Journal of Medicine, 302*(20), 1109–1117.

Prewett, S., Mellard, D. F., Deshler, D. D., Allen, J., Alexander, R., & Stern, A. (2012). Response to intervention in middle schools: Practices and outcomes. *Learning Disabilities Research & Practice, 27,* 136–147.

Puranik, C. S., & Lonigan, C. J. (2012). Name-writing proficiency, not length of name, is associated with preschool children's emergent literacy skills. *Early Childhood Research Quarterly, 27*(2), 284–294.

Puranik, C. S., Petscher, Y., & Lonigan, C. J. (2012). *Dimensionality and reliability of letter writing in 3- to 5-year-old preschool children*. Learning and Individual Differences.

RAND Mathematics Study Panel, Chair, D.L.B. (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education*. Arlington: RAND.

Renaissance Learning (2010). *The foundation of the STAR Assessments*. Wisconsin Rapids: Author.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. (2009). Curriculum-Based Measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47,* 427–469.

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*(4), 546–567.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46*(3), 343–366.

Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77,* 37–59.

Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology, 104*(1), 224–234.

Shapiro, E. S., & Gebhardt, S. N. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review, 41,* 295–305.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24,* 19–35.

Shapiro, E. S., Solari, E., & Petscher, Y. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences, 18,* 316–328.

Sittner Bridges, M., & Catts, H. (2011). The dynamic screening of a phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities, 44*(4), 330–338.

Snow, C., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snowling, M. J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A. M. (2011). Identification of children at risk of dyslexia: The validity of teacher judgments using 'phonic awareness.' *Journal of Research in Reading, 34*(2), 157–170. doi:10.1111/j.1467-9817.2011.01492.x.

Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at risk for reading failure. *Journal of Learning Disabilities, 38*(5), 387–399.

Speece, D. L., Ritchey, K. D., Silverman, R., Schatschneider, C., Walker, C. Y., & Andrusik, K. N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review, 39,* 258–276.

Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary School Journal, 111,* 585–607.

Swanson, H. L., & Howard, C. B. (2005). Children with reading disabilities: Does dynamic assessment help in classification? *Learning Disabilities Quarterly, 28,* 17–34.

Swets, J. A., Tanner, W. P. Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68*(5), 301.

Tackett, K. K., Roberts, G., Baker, S., & Scammaca, N. (2009). *Implementing Response to Intervention: Practices and perspectives from five schools. Frequently asked questions*. Portsmouth: RMC Research Corporation, Center on Instruction.

Teisl, J. T., Mazzocco, M. M., & Myers, G. F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities, 34*(3), 286–293.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measurement in mathematics? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31*(4), 498–513.

Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *The Journal of Special Education, 23*(2), 169–183.

Tolar, T. D., Lederberg, A. R., & Fletcher, J. M. (2012). A structural model of algebra achievement: Computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educational Psychology, 29*(2), 239–266.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator, 22,* 32–41.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15,* 55–64.

Tower, D. M. (1973). A kindergarten screening index to predict reading failure. *Annals of Dyslexia, 23*(1), 90–105.

TPRI. (2013). TPRI Early Reading Assessment. Author: Texas Education Agency & University of Texas System.

VanDerHeyden, A. M. (in press). *Universal screening may not be for everyone: The threshold model as a smarter way to determine risk*. Manuscript accepted for publication in School Psychology Review, 43.

VanDerHeyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children, 77,* 335–350.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30,* 15–31.

VanDerHeyden, A. M., Broussard, C., & Cooley, A. (2006). Further development of measures of early math performance for preschoolers. *Journal of School Psychology, 44,* 533–553.

VanDerHeyden, A. M., Broussard, C., Snyder, P., George, J., Lafleur, S. M., & Williams, C. (2011). Measurement of kindergartners' understanding of early mathematical concepts. *School Psychology Review, 40*(2), 296–306.

Van Waelvelde, H., Hellinckx, T., Peersman, W., & Smits-Engelsman, B. C. (2012). SOS: A screening instrument to identify children with handwriting impairments. *Physical and Occupational Therapy in Pediatrics, 32*(3), 306–319.

Vaughn, S., & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities, 45*(3), 244–256.

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement, 68*(1), 5–24.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2009). *TOSREC: Test of Sentence Reading Efficiency and Comprehension.* Austin: Pro-Ed.

Wagner, R. K., & Compton, D. L. (2011). Dynamic Assessment and Its Implications for RTI Models. *Journal of Learning Disabilities, 44*(4), 311.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41,* 85–120.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24,* 4–12.

Zirkel, P. A., & Thomas, L. B. (2010). State laws for rti: An updated snapshot. *Teaching Exceptional Children, 42*(3), 56–63.

Zumeta, R. O., Compton, D. L., & Fuchs, L. S. (2012). Using word identification fluency to monitor first-grade reading developlemt. *Exceptional Children, 78*(2), 201–220.

The Role of Tier I Mathematics Instruction in Elementary and Middle Schools: Promoting Mathematics Success

Asha K. Jitendra and Danielle N. Dupuis

The purpose of this chapter is to describe effective mathematics instructional programs delivered in general education settings by elementary classroom and middle school mathematics teachers. All of the programs described can be used within tier I of a response-to-intervention (RTI) service delivery model to improve access to core mathematics instruction and promote mathematics success for all students. Although significant advances have been made in developing effective reading programs aimed at improving students' reading ability, the research base on core mathematics (i.e., tier I) programs for elementary and middle grades that could be used in an RTI framework is limited (Clarke et al. 2010; Clarke et al. 2008; Crawford and Ketterlin-Geller 2008; Gersten et al. 2009a). This chapter includes information related to the nature and emphasis of the available instructional programs, the instructional settings in which they have been examined, and how all students, including those struggling in mathematics or at risk for mathematics difficulties, responded to core mathematics instruction. The authors note that while effective tier 1 instruction is critical to meet the instructional needs of a range of learners, paying particular attention to struggling students is important because these students may need subsequent tier 2 or tier 3 instruction, and it would be inappropriate to place students in tier 2 if there is no evidence to suggest they had the opportunity to learn from well-designed tier 1 instruction.

Much effort has been invested in the last two decades on improving mathematics instruction so that all students meet the high standards and expectations as measured by state-administered achievement tests (see NMAP 2008). As such, it is crucial that students at risk for mathematics difficulties, who vary considerably in ability, achievement, and motivation, develop the necessary mathematical knowledge to meet grade-level benchmarks. Within an RTI framework, the standards serve as the primary source of learning objectives for tier 1 instruction and play a key role in determining which students need extra support to prevent mathematics difficulties. Further, there is a need for effective professional tools (e.g., instructional programs) to ensure that tier I instruction adequately supports the development of critical mathematical concepts and skills for all students. This chapter begins with a discussion of critical mathematical concepts and skills to be included in instructional programs and the role of instructional practices in enhancing learning for all students and in preventing mathematics difficulties before they become intractable and present persistent problems for struggling learners. Instructional design features that are supported by research in promoting the mathematical development of struggling students are identified and described. Next, the nature and results of specific mathematics instructional programs in

A. K. Jitendra (✉) · D. N. Dupuis
Department of Educational Psychology, University of Minnesota, 56 E. River Road, 250 Education Science Building, Minneapolis, MN 55455, USA
e-mail: jiten001@umn.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_13

elementary and middle schools are described to understand the instructional conditions that need to be in place to promote mathematics success. Finally, the chapter concludes with a summary of limitations and recommendations for future research.

Promoting Mathematics Success Through Effective Core Instruction: Content and Focus of Instructional Programs and Practices

What elements need to be present for a tier 1 core mathematics instructional program to be successfully implemented in elementary and middle grades? Clarke et al. (2011) identify two key components that are critical for an effective tier 1 core program—(a) a focus on relevant content and (b) use of "research-based instructional design principles" (p. 565). To enhance the learning of all students and prevent serious mathematics difficulties, it is essential to not only examine the general education content and instructional environment (e.g., standards and expectations, materials) but also ensure that research-based practices are implemented to maximize the likelihood that all children have an opportunity to learn.

Reform activities in mathematics education in the first decade of 2000 focused primarily on mathematics curricula, attempting to reduce the number of topics and cover those topics in greater depth (Schmidt and Houang 2007). With the release of the curriculum focal points (2006) by the National Council of Teachers of Mathematics, the focus at each grade level was on critical content areas (e.g., number and operation, geometry, measurement) and instructional standards that emphasized problem-solving, reasoning, and critical thinking. The intent of the curriculum focal points was to circumvent the fragmenting of learning expectations and standards, and to focus curricula and instruction on areas that are foundational to mathematical learning. For example, several researchers and policy documents report that core instruction in the early elementary grades should ensure student development of number sense, which is a foundation for more advanced mathematics encountered in later grades (Milgram and Wu 2005; NMAP 2008). Core instructional programs used as part of tier I instruction should incorporate critical content that develop students' mathematics knowledge. The National Mathematics Advisory Panel (NMAP; 2008) discussed the crucial role of algebra in the overall mathematics curriculum for grades K–8 and provided useful benchmarks for the foundations of algebra (i.e., proficiency with whole numbers, fractions, and particular aspects of geometry and measurement such as similar triangles and properties of two- and three-dimensional shapes) to guide mathematics curricula, instruction, and assessments. With the recent implementation of the common core state standards for mathematics (National Governors Association Center for Best Practices, Council of Chief State School Officers 2010), the ongoing emphasis on providing coverage of a relatively smaller number of topics at each grade level was intentional to ensure robust mathematical understanding. For students struggling with mathematics, the US Department of Education's What Works Clearinghouse (Gersten et al. 2009a) specifically recommended that instructional materials focus on in-depth treatment of whole numbers in elementary grades and on rational numbers (i.e., a number that can be expressed as an integer or a quotient of integers, excluding zero as a denominator) in upper elementary and middle grades. One of the recommendations regarding content to prevent mathematics difficulties was focused "instruction on solving word problems that is based on common underlying structures" (Gersten et al. 2009a, p. 6).

In addition to the content or focus of instructional programs, educators should be cognizant of the instructional practices or instructional design principles that are characteristic of successful programs for all students, including students at risk for mathematics difficulties. The following section provides a summary of the key findings from several meta-analyses on effective instructional practices for students with or at risk for mathematics difficulties (see Jayanthi and Gersten 2011). Results of these syntheses have indicated that instructional programs that incorporated key instructional or curricula design elements benefitted students who struggled with mathematics (Baker et al. 2002; Gersten et al. 2009b; Kroesbergen and Van Luit 2003; Swanson and Hoskyn 1998; Xin and Jitendra 1999). Effective interventions included the following characteristics: (a) providing explicit instruction to teach mathematical concepts and procedures (Baker et al. 2002; Gersten et al. 2009b; Kroesbergen and Van Luit 2003; Swanson and Hoskyn 1998; Xin and Jitendra 1999); (b) teaching students to use heuristics (Gersten et al. 2009b); (c) encouraging students to think aloud while solving a problem (Gersten et al. 2009b); (d) using visual representations of mathematical ideas (Gersten et al. 2009b; Xin and Jitendra 1999); and (e) providing a range of examples and sequencing examples (e.g., concrete to abstract; Gersten et al. 2009b). Research support is unequivocal about the value of explicit and systematic instruction and teaching students to use heuristics (Gersten et al. 2009b; NMAP 2008). Explicit instruction includes "providing models of proficient problem-solving, verbalization of thought processes, guided practice, corrective feedback, and frequent cumulative review" (Gersten et al. 2009a, p. 6). Another promising instructional approach involves increasing student motivation by providing opportunities for solving real-world problems (Gersten et al. 2009b). Finally, other instructional practices found to be effective for students at risk for mathematics difficulties included providing teachers with ongoing feedback about their students' performance using formative assessments, providing feedback only to students, and cross-age tutoring (Gersten et al. 2009b; Baker et al. 2002). Given that tier I instruction is designed to meet the needs of students across the levels of achievement from below grade, at grade, and above grade, understanding the extent to which these instructional practices help to develop student competence in mathematics is critical.

When research-based instructional design elements are incorporated into mathematics instructional programs, a broad range of learners (i.e., students with and without mathematics difficulties) can be expected to achieve important mathematics knowledge and skills. Unfortunately, most tier 1 curricula in general education classrooms do not incorporate critical instructional design elements that are necessary to support the learning of students struggling with mathematics (Bryant et al. 2008; Doabler et al. 2012; Sood and Jitendra 2007). In reviewing studies in the next section, it is important to consider that even well-designed classrooms may not be sufficient to address the instructional needs of all students and that there could still be students who may not respond adequately to core instruction and need more intensive interventions and supports. Furthermore, variability in outcomes for students struggling with mathematics can be attributed to not only the different types of instructional programs but also to differences in the complexity of the content and intensity of instructional programs.

Relevant Research and Evidence of Effectiveness of Tier 1 Instructional Programs

In this section, the focus is on studies examining tier 1 core mathematics instruction, and studies are organized in terms of the intensity of the program described. For the purpose of this chapter, intensity is defined in terms of the total amount of time spent on tier 1 mathematics instruction. The authors arbitrarily categorized studies into three groups—high, medium, and low intensity—on the basis of the total amount of instructional time. Due to space limitations and relevance, studies that used a supplementary tutoring program (e.g., Fuchs et al. 2008, 2009; Jitendra et al. 2013a, b), had a limited content focus (e.g., basic arithmetic facts or procedures; e.g., VanDerHeyden et al. 2012), were delivered for a short duration (e.g., 10 min per day), or linked tier 1 instruction with tier 2 intervention (e.g., VanDerHeyden et al. 2007) were excluded. Furthermore, studies were not included if they used a class-wide mathematics instructional program, but did not discuss the findings for students struggling in mathematics (e.g., Codding et al. 2011; Fuchs et al. 2010; Jitendra et al. 2011). One note of caution about the findings is that not all possible tier 1 studies were reviewed in this chapter, and the studies selected included two studies of an early numeracy program, two studies targeting number sense using a class-wide peer tutoring approach, and seven studies focusing on word problem-solving.

The next section describes each study in terms of the participants and overall school context, as well as the nature and effectiveness of the instructional program. RTI is discussed in terms of the strength of the impact of the instructional program on mathematics outcomes using an effect size (ES) index. Although the studies reviewed reported the effects of the instructional programs for all students, this chapter focuses additionally on students at risk for mathematics difficulties (e.g., scoring low on mathematics tests) identified as such by the authors. ESs were computed for each study using Hedges' *g* as "the difference between pretest–posttest means for the treatment group and the pretest–posttest means for the control group" (see Flynn et al. 2012, p. 24) divided by the pooled posttest standard deviation. Table 1 provides a summary of implications for practice in terms of the resources necessary to implement the interventions.

High Intensity, Explicit Instruction

Two high-intensity studies provided promising findings for a comprehensive mathematics curriculum specially designed for a wide range of learners (Chard et al. 2008; Clarke et al. 2011). Each study evaluated the efficacy of a core kindergarten program, Early Learning in Mathematics (ELM), on the mathematics achievement of kindergartners. The ELM curriculum provides instruction in critical content (i.e., number and operations, measurement, and geometry) and mathematics vocabulary using research-based instructional practices (e.g., explicit and systematic instruction, frequent and cumulative review). The program systematically introduces mathematics topics and develops student understanding of critical concepts. A defining feature of ELM is instruction that is scaffolded, with initial instruction involving extensive modeling and representations of mathematics concepts and gradually removing the supports as students become more and more independent in their work.

Chard et al. (2008) conducted a pilot study to investigate the feasibility and promise of ELM in improving students' mathematics achievement. All 254 kindergarten students (52 % were eligible for free or reduced-cost lunch, 14 % were minority students, and 6 % were English language learners) and their teachers in six treatment and five comparison schools in an urban school district participated in the study. Of the schools in the treatment condition, approximately half of them had implemented an initial version of ELM in the year prior to this study, while the remaining treatment condition schools served as comparison schools in the previous year so that treatment schools in this study had a differential amount of exposure to the ELM program. Although assignment to condition was not random, pretest performance on the measure of mathematics achievement was comparable between treatment and comparison schools. The 100-lesson ELM curriculum was implemented at least 4 days a week, 30 min per session, until all 100 lessons were completed. Students in treatment schools outperformed students in comparison schools on the SESAT-2 (Stanford Early School Achievement Test—Fourth Edition), a standardized measure of achievement ($g=0.32$). Using a mixed-effect ANCOVA (analysis of covariance) model where schools were treated as the unit of analysis, Chard et al. found the difference between treatment and comparison schools to be statistically significant. Chard et al. did not separately report scores for higher- or lower-scoring students in the treatment and comparison groups. However, their analyses found no statistically significant differences between lower- and higher-scoring students, suggesting that ELM was equally effective for students in both lower- and higher-scoring subgroups. As Chard et al. noted, because the school represented the unit of analysis, the design was underpowered with the small number of schools in the study. Nevertheless, it appears that ELM may be promising in improving student learning.

In the next study, Clarke et al. (2011) extended the work of Chard et al. (2008) by investigating the efficacy of the ELM program when compared

Table 1 Summary table of implications for practice

Study	Participants; grades	Intervention and content; duration	Interventionist and training; fidelity of implementation (%)	Measures	Hedges' *g*
High intensity, explicit classroom intervention					
Chard et al. (2008)	Students at risk for mathematics difficulties (scores <25th percentile on TEMA), not diverse participants; KG	Early Learning in Mathematics (ELM—100 lessons on number and operations, measurement, and geometry (critical mathematics vocabulary embedded across the content); on average 4 days per week, 30-min lessons (total time: 3000 min)	General education KG classroom teachers; half-day training prior to implementation of the intervention; >0.80	SESAT-2	NA
Clarke et al. (2011)	Students at risk for mathematics difficulties (scores <40th percentile on TEMA), diverse participants; KG	Early Learning in Mathematics (ELM)—120 lessons on number and operations, measurement, and geometry (critical mathematics vocabulary embedded across the content); 45-min lessons, with supplemental 15-min calendar activities (total time: 7200 min)	General education KG classroom teachers; three 4-h training sessions across the school year; NR	TEMA and CBM total (oral counting, number identification, quantity discrimination)	*TEMA*: ELM vs. C =0.27 *CBM total*: ELM vs. C =0.24
Medium intensity, explicit classroom intervention					
Fuchs et al. (2001)	Low-achieving students (scoring more than 1.5 SDs below the on the SESAT), diverse participants; KG	Peer-assisted learning strategies (PALS)—30 lessons on number concepts, number comparisons, adding/subtracting concepts, and adding/subtracting numerals; on average 2 lessons per week, 20 min each across 15 weeks (total time: 600 min)	General education KG classroom teachers; one 2-h training sessions across the school year; 0.90	SESAT and primary I Stanford Achievement Test	SESAT: PALS> C =0.38 Primary I Stanford: PALS> C =0.30
Fuchs et al. (2002)	Low-achieving students (based on teacher judgment), diverse participants; first grade	PALS)—48 lessons on critical concepts and skills (e.g., recognizing numerals, ordering numerals on a number line, adding/subtracting numerals); on average 3 lessons per week, 30 min each across 16 weeks (total time: 1440 min)	General education KG classroom teachers; one 2-h training sessions across the school year; 0.96	Primary I and II levels of Stanford Achievement Test (aligned and unaligned items)	Primary I and II aligned items: PALS> C =0.19 Primary I and II level unaligned items: PALS> C=0.01

Table 1 (continued)

Study	Participants; grades	Intervention and content; duration	Interventionist and training; fidelity of implementation (%)	Measures	Hedges' *g*
Fuchs et al. (2003a)	Low-achieving students (no criteria reported), diverse participants;third grade	Transfer (Trans) treatment (solution, partial solution + trans, full solution + trans)—26–36 lessons on problem solution rules and word problem-solving involving all four operations and four different problem structures (e.g., shopping list problems, half problems); on average 2 lessons per week, 25–40 min each across 16 weeks (total time: >900 min)	General education third grade classroom teachers and research teachers; NR;0.95–0.99	Researcher-designed word problem tests (pooled immediate and near transfer tests; far transfer test)	*Immediate and near transfer posttests*: Solution vs. C =3.27; Partial solution + trans vs. C =1.33; Full solution + trans vs. C =2.82 *Far transfer posttest*: Solution vs. C =0.61; Partial solution + trans vs. C =0.40; Full solution + trans vs. C =0.97
Fuchs et al. (2003b)	Low-achieving students (based on scores from previous year's district accountability testing), NR; third grade	Transfer treatment (Trans) and trans + self-regulated learning strategies (SRL)—32 lessons on problem solution rules and word problem-solving involving all four operations and four different problem structures (e.g., shopping list problems, half problems); 2 sessions per week across 16 weeks (total time: >900 min)	General education third-grade classroom teachers and research teachers; NR; 95.8	Researcher-designed word problem tests (pooled immediate and near transfer tests; far transfer test)	*Immediate and near transfer posttests*: Trans vs. C =2.29; Trans + SRL vs. C =2.78 *Far transfer posttest*: Trans vs. C =0.95; Trans + SRL vs. C =0.98

Table 1 (continued)

Study	Participants; grades	Intervention and content; duration	Interventionist and training; fidelity of implementation (%)	Measures	Hedges' *g*
Fuchs et al. (2004a)	Low-achieving students (scores <25th percentile on problem-solving pretest), diverse participants; third grade	Schema-based transfer instruction (SBTI) and expanded SBTI—34 lessons on general math problem-solving strategies word problems involving all four operations and four different problem structures (i.e., shopping list, half, buying bags, and pictograph problems); 2 lessons per week, 25–40 min each across 16 weeks (total time: >1000 min)	General education third-grade classroom teachers; NR; 95.9	Researcher designed word problem tests (pooled transfer one, two, and three tests; Transfer four test)	*Transfer one, two, and three posttests*: SBTI vs. C = 3.31; Expanded SBTI vs. C = 3.43 *Transfer 4 posttest:* SBTI vs. C = 1.40; Expanded SBTI vs. C = 2.15
Fuchs et al. (2004b)	Low-achieving students (scores <25th percentile on problem-solving pretest), diverse participants; third grade	Schema-broadening instruction (SBI) and SBI plus sorting (SBI+S)—30–32 lessons on general math problem-solving strategies and word problems involving all four operations and four different problem structures (i.e., shopping list, half, buying bags, and pictograph problems); 2 lessons per week, 30–40 min each across 16 weeks (total time: 980–1040 min)	General education third-grade classroom teachers; NR; 95.5	Researcher-designed word problem tests (pooled immediate and near transfer tests; far transfer test)	*Immediate and near transfer posttests:* SBI vs. C = 6.54; SBI+S vs. C = 4.65 *Far transfer posttest:* SBI vs. C = 1.84; SBI+S vs. C = 1.87
Jitendra et al. (2007)	At-risk students (LD, title 1, ELL), diverse participants; third grade	Schema-based instruction (SBI)—36 lessons on one-step and two-step addition and subtraction word problems involving *change, group, and compare* problem types and 3 review lessons; 39 lessons, 25 min daily for 12 weeks (total time: 1500 min)	General education third-grade classroom teachers and a special education teacher; two 2-h training sessions; 96.3	Researcher-designed word problem-solving (WPS) tests—immediate and delayed (6 weeks later); PSSA	*Immediate posttest*: SBI vs. C = 0.41 *Delayed posttest*: SBI vs. C = 0.85 *PSSA*: SBI vs. C = 0.80

Table 1 (continued)

Study	Participants; grades	Intervention and content; duration	Interventionist and training; fidelity of implementation (%)	Measures	Hedges' *g*
Low intensity, explicit classroom intervention					
Jitendra et al. 2009	Low-achieving students (based on grades in mathematics from previous school year and scores on mathematics subtests of SAT-10), diverse participants; seventh grade	Schema-based instruction (SBI)— – 10 lessons on ratio and proportion; 10 lessons, 40 min each (total time: 400 min)	General education seventh-grade math teachers; 1-day training session; 80.0	Problem-solving (ratio and proportion) tests (items derived from TIMSS, NAEP, and state assessments) and PSSA	*Immediate posttest:* SBI vs. C = 0.006 *Delayed posttest* (*4 months later*): SBI vs. C = 0.26 *PSSA*: SBI vs. C = −0.22)
Jitendra and Star 2012	Low-achieving students (based on grades in mathematics from previous school year), diverse participants; seventh grade	Schema-based instruction (SBI)— 6 lessons on percent and 3 cumulative review lessons; 9 sessions, 40 min each (total time: 360 min)	General education seventh-grade math teachers; 1 half-day training session; 76.0	Problem-solving (percent) and transfer tests (items derived from TIMSS, NAEP, and state assessments)	*Immediate posttest*: SBI vs. C = −0.39 *Transfer test:* SBI vs. C = −0.10

C control or contrast, *CBM* curriculum-based measurement, *ELL* English language learner, *LD* learning disabilities, *PSSA* Pennsylvania System of School Assessment, *SESAT-2* Stanford Early School Achievement Test-Fourth Edition, *TEMA* Test of Mathematics Ability-Third Edition, *NAEP* National Assessment of Educational Progress, *NR* not reported, *TIMSS* Third International Mathematics and Science Study, *SAT−10* Stanford Achievement Test—Tenth Edition

to standard district practice on the mathematics achievement of all students, including students at risk for mathematics difficulties, using a rigorous randomized controlled trial. Blocking by school, 64 kindergarten classrooms (matched on full or partial day) were randomly assigned to treatment (n=34) or control (n=30) conditions. Sixty-five teachers (one classroom included two teachers) and their students participated (56% were eligible for free or reduced-price lunch, 51% were minority students, 38% were English language learners, and 8% were receiving special education services). A cut score of the 40th percentile on the Test of Mathematics Abilities—Third Edition (TEMA-3) was used to identify students as at risk or not at risk for mathematics difficulties. The ELM curriculum was implemented daily for 60 min.

Regarding students not at risk for mathematics difficulties, results indicated no statistically significant differences between treatment and control students on the TEMA-3 (g=0.03) or on the early numeracy curriculum-based measurement (EN-CBM; g=0.04). In contrast, for students identified as at risk for mathematics difficulties, statistically significant differences favored students in the treatment classrooms compared to students in the control classrooms on the TEMA-3 (g=0.27) and on the EN-CBM (g=0.24). On both measures, a comparison of the performance of students at risk for mathematics difficulties in relation to the performance of their peers not at risk for mathematics difficulties in ELM and control classrooms showed that the gains made by ELM students at risk for mathematics difficulties were greater than the gains made by control students at risk for mathematics difficulties. Furthermore, it is encouraging that "9.2 % more students in ELM classrooms transitioned from the at-risk to not-at-risk categories than in control classrooms, a difference that was statistically significant" (Clarke et al. 2011, p. 577). In sum, although the effects for both ELM studies (Chard et al. 2008; Clarke et al. 2011) are small and modest, these results provide preliminary evidence regarding the effectiveness of the ELM program to improve student mathematics achievement, particularly for students at risk for mathematics difficulties. At the same time, it might be argued that there is insufficient evidence for ELM benefitting students across all levels of achievement and that ELM might be a better tier 2 intervention based on positive effects for low-performing students only as seen in the Clarke et al. (2011) study. On the one hand, tier 1 instruction that incorporates research-based instructional practices may be effective for students performing at grade and above grade levels only when instruction is appropriately differentiated such that the content is sufficiently challenging for these students. On the other hand, ELM as a tier 1 intervention may be necessary for students at risk for mathematics difficulties to benefit from tier 2 instruction. The authors recently tested the added value of tier 2 tutoring intervention over and above tier I instruction with the ELM curriculum. Positive benefits were seen for tutored students compared to control students receiving only tier I instruction with the ELM curriculum (Clarke et al. 2013).

Medium Intensity, Explicit Instruction

Peer-Assisted Learning Strategies (PALS) To understand whether peer-mediated instruction can promote mathematics success for students, two of the most recent and related studies from the peer-mediated, class-wide intervention research were examined. The rationale for focusing on only two studies is based on space limitations, and also because peer-mediated instruction is less common than teacher-directed instruction in general education classrooms (Fuchs et al. 2001, 2002). In each study, Fuchs and colleagues assessed the effects of peer-assisted learning strategies (PALS), a version of class-wide peer tutoring, on children's mathematical development (see Fuchs et al. 1995 for a review of PALS). The studies contrasted PALS with conventional classroom mathematics instruction, in which teachers relied on the instructional activities in their district-adopted mathematics textbook to teach critical mathematics content. During PALS, teachers selected critical content and skills from the district's mathematics textbook and replaced parts of their mathematics instruction (teacher-directed instruction and student workgroup activities) with peer-mediated class-wide instruction by pairing higher- and lower-achieving students to practice applying the learned content (e.g., number sense) and skills.

Fuchs et al. (2001) randomly assigned 20 kindergarten teachers in three title 1 and two nontitle 1 urban schools to either PALS or a no-PALS condition. In total, 153 students participated (44 % were eligible for free or reduced-price lunch, 63 % were minority students, 2 % were English learners, and 10 % were receiving special education services). The PALS curriculum was adapted from the RightStart program (Griffin et al. 1994) that focuses on number sense. Treatment teachers implemented PALS twice a week for 20 min across 15 weeks. To understand whether or not PALS was effective for all students, Fuchs et al. classified students into three achievement groups (high, average, low) using scores from the SESAT. On the SESAT, both average- and low-achieving students in the PALS condition outperformed their no-PALS peers, with ESs of 0.44 and 0.38, respectively. In contrast, the effectiveness of PALS for high-achieving students compared to high-achieving students in the no-PALS condition was questionable ($g = -0.16$). The authors argued that, perhaps, a number sense curriculum was not sufficiently challenging to address these students' needs. Relatedly, Fuchs et al. also collected data on the primary 1 level of the Stanford Achievement Test (SAT) to address the issue that high-achieving students would experience a ceiling effect on the SESAT; additionally, the primary 1 is not as closely aligned with the instructional content taught in PALS as is the SESAT. As such, the primary 1 provides a more distal measure of student achievement than the SESAT. On the primary 1 measure, both high- and low-achieving students in the PALS condition outperformed their no-PALS peers, with ESs of 0.66 and 0.30, respectively. Given the earlier explanation that high-achieving students in the PALS condition might have already mastered the majority of the content in the PALS curriculum, why did these children perform better than their high-achieving no-PALS peers on this measure? The authors attributed the positive, moderate-to-large effect to the strategy of explaining to others and reported "pairing children did ensure routine opportunities for HA students to work together on more complex skills using high number sets" (p. 507). Surprisingly, the ES for average-achieving students on the primary 1 measure was small ($g = -0.21$), favoring the no-PALS condition.

In the next study, Fuchs et al. (2002) examined whether PALS would be efficacious for first-grade students in improving their mathematics knowledge. Blocking by school, 20 classrooms were randomly assigned to PALS or no-PALS conditions. All classrooms were part of a single urban district, where eight classrooms in each condition were located in title I schools and the other two classrooms (in each condition) were in nontitle I schools. The final sample included 327 students (62% were eligible for free or reduced-price lunch, 75% were minority students, 31% were English language learners, and 6% were receiving special education services). PALS teachers led three 30-min sessions a week for 16 weeks. The intensity of PALS in this study was more than twice the total time that PALS was implemented in Fuchs et al. (2001). At the beginning of the study, students were classified as high, average, or low performing based on teacher judgments of students' classroom performance. To assess student learning, the primary 1 and primary 2 levels of the SAT were administered before and after the intervention was implemented. In addition, each item on the primary 1 and primary 2 was categorized as being aligned with PALS or not aligned with PALS. Results were examined separately for each of the three achievement groups and for the aligned and unaligned items. For the items aligned with PALS, effects were negligible even though they favored the PALS condition for high-, average-, and low-performing students, with ESs of 0.15, 0.16, and 0.19, respectively. In contrast and consistent with the authors' hypotheses, no statistically significant between-condition (PALS and no PALS) differences were found for items that were not aligned with PALS. ESs for high-, average-, and low-performing students were −0.05, −0.07, and 0.01, respectively. In sum, findings about the effectiveness of PALS as a tier 1 intervention are somewhat mixed based on the two studies that provided focused instruction on number sense using a PALS curriculum adapted from the RightStart program (Griffin et al. 1994). It may be that additional studies about the effectiveness of PALS as a tier I intervention are needed in which instruction is appropriately differentiated for students across all levels of achievement.

Schema-Based/Schema-Broadening Instruction To understand the effectiveness of instructional strategies that build student capacity to solve word problems by focusing on common underlying structures, findings from two teams of researchers led by Fuchs and Jitendra are examined. Fuchs and colleagues' instructional approach to solving word problems emphasizes schema-broadening instruction and is defined by an explicit focus on transfer in the instructional design. A schema is a specific knowledge structure that individuals use to categorize problem types into groups requiring similar solutions (Gick and Holyoak 1983). All tier 1 studies of schema-broadening instruction by Fuchs and colleagues included the following problem types: shopping list problems, half problems, buying bag problems, and pictograph problems. In an ongoing series of investigations, Fuchs and colleagues have examined the effects of schema-broadening instruction to promote "transfer to problems with unexpected features within the taught problem types (e.g., irrelevant information, relevant information presented outside the narrative in figures or tables, presentation of problems in real-life context)" (Schumacher and Fuchs 2012, p. 611). Transfer occurs when the learner recognizes that novel problems, even though different in certain features, are related to previously solved problems. The authors introduced the concept of transfer (i.e., *to move*) using several examples from daily life and mathematics (e.g., *move* from adding two-digit horizontal problems to solving two-digit vertical problems). In all of these studies, schema-broadening instruction was contrasted with conventional classroom problem-solving instruction (i.e., teacher-delivered instruction on problem–solution rules, which was based on the district-adopted mathematics textbook and included the four target problem types that were the focus of instruction in each of the four studies described herein).

Fuchs et al. (2003a) randomly assigned 24 third-grade teachers in six urban schools to one of four conditions (three treatment and a control condition); each condition contained six teachers and conditions were stratified across schools. In total, 375 students provided data at pretest and posttest (45 % were eligible for free or reduced-price lunch, 56 % were minority students, 4 % were English language learners, and 6 % were receiving special education services). The 16-week intervention took one of three forms (i.e., the three treatment conditions): teachers and research assistants provided (a) solution instruction (26 lessons) that taught rules for problem-solving and sorting problems into specific problem types that require the same solution methods, (b) partial solution instruction plus transfer instruction (26 lessons), and (c) full solution instruction plus transfer instruction (36 lessons). Transfer instruction focused on explicitly teaching the concept of transfer, highlighting superficial problem features (i.e., different format, different key word, additional or different question, and problem scope—problem is placed within a larger problem-solving context) that can modify a problem but not the structure or solution, and understanding that novel problems could incorporate these superficial features and yet represent a familiar problem structure or solution. Word problem-solving instruction occurred two times a week for 25-40 min across 16 weeks.

Student learning was assessed using three tests of transfer—immediate (four problem types with novel cover stories), near (four problem types with novel cover stories and with one superficial feature varied per problem), and far (all four problem types "embedded in a real-life problem-solving context, with all four superficial features varied and additional elements of novelty" Fuchs et al. 2003a, p. 299). Students were classified as high, average, or low achieving, and results were analyzed separately for each transfer measure and students' achievement status. Students in each treatment condition significantly outperformed control students across the three measures and in each achievement group. Across the three treatment conditions, ESs for students identified as high, average, and low achieving on the measure of immediate transfer were large and ranged from 1.29 to 3.26, 1.92 to 5.45, and 1.36 to 3.45, respectively. On the measure of near transfer, although the ESs were large, they were generally smaller than the ESs for the measure of immediate transfer. ESs for high-, average-, and low-achieving students ranged from 1.05 to 2.56, 0.87 to 1.93, and 1.30 to 3.10, respectively. Finally, although ESs on the measure of far transfer were considerably smaller than on the measures of near and immediate transfer, the effect sizes can be considered moderate to large. ESs for high-, average-, and low-achieving students ranged from 0.40 to 1.41, 0.22 to 0.67, and 0.40 to 0.97, respectively.

In a follow-up study, Fuchs et al. (2003b) evaluated the contribution of self-regulated learning strategies (SRL) incorporated into problem-solving transfer instruction. Self-regulation learning strategies consisted of goal setting and self-monitoring. Students set goals for their performance, scored performance on problem-solving tasks, and graphed their scores. Twenty-four third-grade teachers from six urban schools were randomly assigned to one of three conditions (two treatment or control conditions); 395 students participated. Treatment teachers and research assistants either provided (a) transfer instruction that was similar to the Fuchs et al. (2003a) solution plus transfer condition or (b) transfer plus SRL that incorporated goal setting and self-evaluation in addition to transfer instruction. Instruction occurred two times a week, across 16 weeks. As in Fuchs et al. (2003a), students were assessed on three measures of transfer, and students were classified into one of three achievement groups (high, average, low). Students in both treatment conditions outperformed control students on all three measures of transfer. Across measures, ESs for the high-, average-, and low-achieving students ranged from 0.43 to 4.41, 0.57 to 5.65, and 0.95 to 3.17, respectively. As might be expected, in all cases, ESs were largest for the measure of immediate transfer and smallest for the measure of far transfer.

Regarding the effect of SRL, moderate-to-large effects favored the transfer plus SRL condition over the transfer-only condition on the measures of immediate and near transfer; ESs for high-, average-, and low-achieving students were 1.00 and 0.92, 0.81 and 0.51, and 0.31 and 0.32, respectively. On the measure of far transfer, small effects favored the transfer plus SRL condition over the transfer-only condition for high-, average-, and low-achieving students with ESs of 0.23, 0.13, and 0.27, respectively. Together, these results suggest that incorporating goal setting and self-monitoring into transfer instruction produces positive outcomes for all students.

The third study (Fuchs et al. 2004b) examined the added value of practice in sorting problems into specific problem types when embedded in the Fuchs et al. (2003a) transfer instruction. Twenty-four third-grade teachers from six urban schools were randomly assigned to one of three conditions (two treatment or control); 366 students provided pretest and posttest data (37% were eligible for free or reduced-price lunch, 38% were minority students, and 8% were English language learners). Treatment teachers and research assistants either provided (a) schema-based transfer instruction (see Fuchs et al. 2003a) or (b) schema-based transfer instruction plus instruction on sorting problems into specific problem types. Word problem-solving instruction occurred two times a week for 30-40 min across 16 weeks. Student learning was assessed using three measures of transfer (immediate, near, and far). Using pretest scores on the test of immediate transfer, students were categorized into one of three achievement groups (high, average, low), and results were examined separately for each group.

As in the Fuchs studies described above, large and statistically significant effects were seen on all measures for students in all achievement groups, favoring the treatment conditions over the control condition. Across measures and both treatment–control comparisons, the ESs for high-, average-, and low-achieving students ranged from 0.76 to 4.66, 0.91 to 4.21, and 1.84 to 8.02, respectively. However, results comparing the treatment conditions were more mixed. On the measure of immediate transfer, although there was no effect for adding sorting instruction to schema-based transfer instruction for high- ($g=-0.08$) and low-achieving ($g=0.03$) students, a moderate, positive effect was found for average-achieving students ($g=0.54$). In contrast, on the measure of near transfer, results revealed moderate, negative effects for high- ($g=-0.64$) and low-achieving ($g=-0.59$) students, but no effect for average-achieving students ($g=0.06$). Finally, on the measure of far transfer, results revealed small negative effects for high- ($g=-0.28$) and low-achieving ($g=-0.20$) students, and a negligible effect for average-achieving students ($g=0.10$). Together, these results suggest that both high- and low-achieving students did not benefit from sorting instruction. In fact, the inclusion of sorting instruction for these students appears to have inhibited learning, particularly on the measure of near transfer. For average-achieving students, the results were more positive; however, significant effects were found only on the measure of immediate transfer. In sum, significant positive effects favoring the inclusion of sorting instruction were found only for average-achieving students on one measure, suggesting little to no benefit of including sorting instruction in schema-based transfer instruction for low- and high-achieving students.

In a fourth study, Fuchs et al. (2004a) estimated the value of an expanded schema-based transfer instruction that included three additional real-life transfer features—"irrelevant information, combining of problem types, and mixing of superficial features" (p. 423) that are considered more challenging than the transfer features (different format, different vocabulary, different question) in previous schema-based transfer instruction studies. Blocking by school, 24 third-grade teachers from seven urban schools were randomly assigned to one of three conditions (two treatment or control); 351 students participated (48 % were eligible for free or reduced-price lunch, 55 % were minority students, 4 % were English language learners, and 8 % were receiving special education services). The treatment conditions included either (a) schema-based transfer instruction (see Fuchs et al. 2003a) or (b) expanded schema-based transfer instruction that included three additional transfer features. Word problem-solving instruction was conducted for 16 weeks, and included 34 lessons with each lesson occurring for 25–40 min. To assess student learning, four measures of transfer problem-solving, all of which included novel problems, were administered at pretest and posttest. The measures were labeled transfer 1 through transfer 4, with greater numbers representing greater transfer distance. As in previously described studies, students were identified as high, average, or low achieving, and results were analyzed separately for each measure and for each student achievement group.

When comparing the treatment conditions to the control condition, large positive effects favoring the treatment conditions were found on all measures for students in all achievement groups. Across measures and treatment–control comparisons, the ESs for high-, average-, and low-achieving students ranged from 0.72 to 5.91, 0.78 to 7.93, and 1.40 to 4.67, respectively. In all cases, ESs were largest on the measure of nearest transfer (i.e., transfer 1) and smallest on the measure of farthest transfer (i.e., transfer 4). In contrast, while the results comparing expanded schema-based transfer instruction to schema-based transfer instruction were mixed across student achievement groups on transfer 1 and transfer 2, moderate-to-large, positive effects ($g = 0.55–1.28$) favored expanded schema-based transfer instruction on the transfer 3 and transfer 4 measures for all student achievement groups.

To understand the short- and long-term effects of schema-based instruction (SBI) in enhancing students' word problem-solving performance and mathematics achievement, findings from Jitendra et al. (2007) are examined. SBI is a multicomponent intervention based on schema theories of cognitive psychology, research on expert problem solvers, and research regarding effective instructional practices (e.g., explicit instruction) for students at risk for mathematics difficulties. Specifically, SBI for elementary grades includes four critical elements: (a) priming the underlying problem structure, (b) using visual representations (schematic diagrams), (c) explicitly teaching of problem-solving heuristics, and (d) providing instruction in metacognitive strategy knowledge. Jitendra et al. (2007) randomly assigned 88 third-grade students (49 % were eligible for free or reduced-price lunch, 50 % were minority students, 5 % were English language learners, and 13 % were receiving title 1 services) and their six teachers in one of the lowest-achieving schools in an urban school district to one of two conditions: SBI or comparison (general strategy instruction, GSI). Students in the GSI condition were taught a heuristic to (a) understand the problem, (b) plan to solve the problem, (c) solve the problem, and (d) look back or check. Problem-solving strategies such as using objects, acting out the problem or drawing a diagram, choosing an operation—writing a number sentence, and using data from a graph or table were incorporated in the planning step of the problem-solving heuristic. Students in both conditions received word problem-solving instruction on one-step and two-step addition problem structures (i.e., change, group, and compare) 5 days a week for 25 min in addition to 25 min of daily core mathematics instruction across 12 weeks of the study. To assess student learning, a measure of word problem-solving was administered at pretest, posttest, and delayed posttest (6 weeks following the end of the intervention). In addition, data were collected on a state-administered test of mathematics achievement. Across measures, findings are presented separately for all students and a subgroup of at-risk students (i.e., learning disabilities, math title 1, English language learners) identified as such by the authors.

On the measure of word problem-solving, results revealed small-to-moderate, positive effects favoring SBI at posttest and moderate-to-large, positive effects favoring SBI on the delayed posttest. On the immediate posttest, ESs for not-at-risk and at-risk students were 0.39 and 0.41, respectively; on the delayed posttest, ESs for not-at-risk and at-risk students were 0.51 and 0.85, respectively. Similarly, results revealed moderate-to-large, positive ESs favoring SBI for both not-at-risk ($g=0.46$) at-risk ($g=0.80$) students on the state-administered test of mathematics achievement. Together, these results provide evidence supporting the use of SBI (i.e., priming the underlying problem structure using schematic diagrams, explicitly teaching a problem-solving heuristic, emphasizing metacognitive strategy use) for all students, including those at risk for mathematics difficulties.

Low Intensity, Explicit Instruction

Two final studies by Jitendra and colleagues provide findings about the effectiveness of SBI for seventh-grade students. These two pilot studies were conducted in an urban, middle school. The first study (e.g., Jitendra et al. 2009) focused on ratio and proportion problem-solving and occurred in early January; the second study (Jitendra and Star 2012) focused on percent problem-solving and was conducted in May of the same year. In both studies, Jitendra and colleagues reviewed the district-adopted mathematics textbook, identified specific concepts and problem-solving skills, and mapped the relevant topics to the SBI units on ratio, proportion, and percent. Jitendra et al. (2009) examined the potential impact of a 10-day intervention (SBI) for 148 students (42 % were eligible for free or reduced-price lunch, 54 % were minority students, 3 % were English language learners, and 10 % were receiving special education services) from eight seventh-grade mathematics classrooms. Blocking by classroom ability level (high, average, and low) based on grades in mathematics from the previous school year and scores on mathematics subtests of SAT-10, classrooms were randomly assigned to either SBI or a "business-as-usual" control condition that received the same amount of instruction on the same topics (i.e., ratio and proportion). Mathematics teachers provided all instruction that occurred during the regularly scheduled mathematics instructional period, 5 days a week for 40 min across 10 school days. In addition to priming the underlying problem structure, using visual representations, explicitly teaching problem-solving heuristics, and providing instruction in metacognitive strategy use, SBI emphasized multiple solution strategies (cross multiplication, unit rate, and equivalent fractions). To assess student learning, data were collected on a measure of problem-solving immediately following completion of the intervention (posttest), and again 4 months later to assess maintenance effects (delayed posttest). In addition, data were collected on a state-administered test of mathematics achievement. Students were categorized as low, average, or high achieving, and the results are presented separately for each student achievement group.

On the test of problem-solving, moderate-to-large ESs were found for both average- and high-achieving students, favoring SBI on the immediate and delayed posttests. For average-achieving students, the ESs on the immediate posttest and delayed posttest were 0.86 and 0.47, respectively; for high-achieving students, the ESs for the immediate and delayed posttests were 0.75 and 1.31, respectively. On the test of mathematics achievement, results revealed no effect for average-achieving students ($g=-0.03$) and a moderate effect favoring SBI for high-achieving students ($g=0.52$). In contrast, the findings for low-achieving students were mixed. No effect was found for SBI at posttest ($g=0.01$) for low-achieving students, whereas a small effect favoring SBI was found at delayed posttest ($g=0.26$) for low-achieving students. On the test of mathematics achievement, a small effect ($g=-0.22$) favored the control condition for low-achieving students. As Jitendra et al. (2009) suggested "the value of integrating metacognitive strategy knowledge as an instructional feature in schema-based instruction, particularly using schematic diagrams to represent information, may not have been realized in the short-term (10-day) intervention for low-ability students" (p. 260). This feature of the intervention and the fact that these students' reported difficulty in mastering the two conceptually based strategies (unit rate, equivalent fractions) may explain their persistent problem-solving difficulties, suggesting the need for more practice, time, and scaffolding of instruction to positively impact learning.

In the second study (Jitendra and Star 2012), data were collected from four (two high-ability and two low-ability classrooms) of the eight seventh-grade mathematics classrooms in the Jitendra et al. (2009) study. A total of 70 students (36% were eligible for free or reduced-price lunch, 59% were minority students, 4% were English language learners, and 7% were receiving special education services) participated in the study that examined the effects of SBI on student learning of percent word problems. Teachers led daily 40-min sessions for 9 days in SBI and control classrooms. The content focused on percent problem-solving, which is a difficult topic for many middle school students (see Parker and Leinhardt 1995). Results revealed that while high-achieving students in the SBI condition statistically outperformed high-achieving students in the control condition on percent problem-solving posttest ($g=0.97$), no statistically significant differences were found for low-achieving students ($g=-0.39$); and a small-to-medium ES favored students in the control condition. On the transfer test, no statistically significant differences were found for either high- ($g=-0.01$) or low- ($g=-0.10$) achieving students. Jitendra and Star (2012) reported that low-achieving students did not respond successfully to the SBI intervention in this study. Unlike previous studies (e.g., Jitendra et al. 2007) that implemented SBI for 12 weeks on average, the 9-day intervention was likely not sufficient for low achievers, who may have needed more time and support to recognize the underlying problem structure and "show gains in flexible knowledge of procedures for solving a wide range of problems" (p. 157).

Limitations and Implications for Future Research

The studies reviewed suggest that well-designed, core mathematics instructional programs applied within an RTI framework could address the diverse needs of a range of learners when instruction is differentiated in ways that not only is the content appropriately matched to the instructional needs of students performing below grade level but also the content is sufficiently challenging for students above grade level. Common features of instructional programs in the studies reviewed were explicit and systematic instruction, scaffolded instruction with teachers or peers first modeling followed by teacher-guided practice with corrective feedback until students are able to work independently, and development of student initial understanding of critical concepts and procedures using a range of examples. Several programs also included other features such as teaching to mastery (e.g., Clarke et al. 2011), providing adequate opportunities in terms of frequent cumulative review to learn the concepts and skills (e.g., Clarke et al. 2011; Fuchs et al. 2003a, b; Jitendra et al. 2007), having students verbalize their thinking (e.g., Clarke et al. 2011; Fuchs et al. 2003b; Jitendra et al. 2007, 2009; Jitendra and Star 2012) using mathematical representations (e.g., Clarke et al. 2011; Jitendra et al. 2007, 2009; Jitendra and Star 2012), teaching the use of heuristics to solve word problems (Jitendra et al. 2007, 2009; Jitendra and Star 2012). It is important that adequate progress monitoring and universal screening measures are used during and at the end of tier 1 intervention to detect children who are still struggling and provide them with tier 2 or tier 3 intervention that incorporates these key instructional features.

An important finding across the studies reviewed is that differences in responsiveness to instruction for students at risk for mathematics difficulties appeared to be associated with variations in the intensity of the instructional programs and content. In several of the medium-intensity studies, moderate-to-large ESs for students at risk for mathematics difficulties (MD) favored schema-based/schema-broadening instruction that focused on word problem-solving; the effects for peer-mediated instruction that focused on number sense were mixed and ranged from no effect to small-to-moderate effects. In contrast, the effects for students at risk for mathematics difficulties in the high-intensity studies that focused on the entire kindergarten content (e.g., number sense, measurement, geometry) were small, and the effects were mixed for students at risk for mathematics difficulties in the low-intensity studies that addressed complex content (e.g., ratio and proportions).

While the research base described above provides important evidence about the effectiveness of various core mathematics instructional programs in improving the mathematics success of a range of learners, there are a number of limitations of the current research that suggest important avenues for future research. First, little is known about how best to support the development of advanced mathematics (e.g., proportional reasoning, algebra, geometry). The majority of studies examined elementary school student populations and content, with only two studies conducted in middle schools and none at the high school level. Of note, none of the high- or medium-intensity interventions that were reviewed examined the effectiveness of instructional programs for students at the secondary grades. A lack of research in secondary settings is of concern for two reasons. First, it is critical that instructional programs implemented within an RTI framework are evidence based, and currently, there is a deficit in the evidence about effective core mathematics programs for the secondary grades. Second, the complexity of the instructional content covered in secondary school (e.g., ratios and proportions, probability) is greater than that covered in elementary school (e.g., whole numbers), which may limit the generalizability of findings about programs tested in elementary grades to the secondary grades. As such, more research is needed about the effectiveness of instructional programs designed to address more advanced content (e.g., rational numbers, algebra, geometry).

A second limitation of the current research involves fidelity of implementation. The National Center on Response to Intervention (NCRI) defines fidelity of implementation as "the delivery of content and instructional strategies in the way in which they were designed and intended to be delivered: accurately and consistently. Although interventions are aimed at learners, fidelity measures focus on the individuals who provide the instruction" (NCRI n.d.). It is critical that researchers conducting studies on the effectiveness of instructional programs develop valid and reliable fidelity measures that can be used to not only assess the extent to which programs are implemented as intended but also the extent to which fidelity affects program effectiveness. In all the studies reviewed, fidelity measures were used to assess implementation; however, none of the studies directly related the fidelity data to outcomes. As such, these studies only provide descriptions of the extent to which programs were implemented with fidelity, and fail to address the extent to which fidelity affected program effectiveness in promoting mathematics success for all students.

A related limitation involves the professional training of teachers, specifically the quality and intensity of the professional development provided. In the studies described above, teachers received anywhere from 2 to 12 h of professional development. For example, in three studies, teachers participated in a single 2-h workshop, while in another study teachers attended three half-day professional development sessions throughout the duration of the study. What is less known from the studies reviewed is how much training is necessary for teachers to implement these programs with fidelity to improve student learning.

Finally, a methodological limitation that is inherent to all classroom-based research is the nesting of students within classrooms and classrooms within schools. Failure to account for this nesting and the resulting dependency between units of analysis produces biased estimates of relevant effects. In the studies considered, the respective researchers chose a variety of approaches designed to address the issues (e.g., power, independence) associated with nested designs. Unfortunately, in most of the studies, the authors did not utilize the most advanced statistical methods available (e.g., multilevel modeling); the result in most cases was a loss of statistical power. A related issue involves disagreement between the unit of assignment and the unit of inference. For example, in most of the studies described above, the instructional program of interest was assigned to whole classrooms and in some cases whole schools; however, the unit of inference in all cases was students. This disconnect between the unit of assignment and the unit of inference leads to what is known as the "ecological fallacy." The ecological fallacy is a logical fallacy specific to the interpretation of statistical data and indicates that correlations between variables at the aggregate level (e.g., classrooms) are not equal to correlations (between the same variables) at the individual level (e.g., students). This means that fitting a regression model to group-level data will not necessarily produce the same effects as fitting that same regression model to individual-level data.

In conclusion, based on the studies summarized in this chapter, several lessons were learned about the instructional conditions that need to be in place to insure that all students, including student at risk for mathematics difficulties, are successful in meeting the requirements for mathematics content involving whole numbers. However, less is known about what is needed to prepare students in meeting the standards and expectations for advanced mathematics in late elementary, middle, and high school.

References

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103,* 51–73.

Bryant, B., Bryant, D. P., Kethley, C., Kim, S. A., Pool, C., & You-Jin, S. (2008). Preventing mathematics difficulties in the primary grades: The critical features of instruction in textbooks as part of the equation. *Learning Disabilities Quarterly, 31,* 21–35.

Chard, D. J., Baker, S. K., Clarke, B., Jungjohann, K., Davis, K., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disability Quarterly, 31,* 11–20.

Clarke, B., Baker, S. K., & Chard, D. (2008). Best practices in mathematics assessment and intervention with elementary students. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 453–464). Bethesda: National Association of School Psychologists.

Clarke, B., Gersten, R., & Newman-Gonchar, R. (2010). RTI in mathematics: Beginnings of a knowledge base. In T. A. Glover & S. Vaughn (Eds.), *The promise of response to intervention: Evaluating current science and practice* (pp. 187–202). New York: Guilford.

Clarke, B., Smolkowski, K., Baker, S. K., Hank, F., Doabler, C. T., & Chard, D. J. (2011). The impact of a comprehensive tier I core kindergarten program on the achievement of students at risk in mathematics. *The Elementary School Journal, 111,* 561–584.

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., & Fien, H. (2013, February). *Evaluating the value of a tier 2 intervention in kindergarten mathematics*. Paper presented at the Pacific Coast Research Conference. San Diego, CA.

Codding, R. S., Chan-Iannetta, L., George, S., Ferreira, K., & Volpe, R. (2011). Early number skills: Examining the effects of class-wide interventions on kindergarten performance. *School Psychology Quarterly, 26,* 85–96.

Crawford, L., & Ketterlin-Geller, L. R. (2008). Improving math programming for students at risk: Introduction to the special topic issue. *Remedial and Special Education,* 29, 5–8.

Doabler, C. T., Fien, H., Nelson-Walker, N. J., & Baker, S. K. (2012). Evaluating three elementary mathematics programs for presence of eight research-based instructional design principles. *Learning Disability Quarterly, 35,* 200–211.

Flynn, L. J., Zheng, X., & Swanson, H. L. (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research & Practice, 27,* 21–32.

Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24,* 604–20.

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal, 101,* 495–510.

Fuchs, L. S., Fuchs, D., Yazdian, L., & Powell, S. R. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review, 31,* 569–583.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Jancek, D. (2003a). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95,* 293–305.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K. (2003b). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95,* 306–315.

Fuchs, L. S., Fuchs, D., Finelli, R., Courey, S. J., & Hamlett, C. L. (2004a). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal, 41,* 419–445.

Fuchs, L. S., Fuchs, D., Prentice, K., Hamlett, C. L., Finelli, R., & Courey, S. J. (2004b). Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology, 96,* 635–647.

Fuchs, L. S., Fuchs, D., Craddock, C., Schatschneider, C., Hollenbeck, K. N., & Hamlett, C. L. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology, 100,* 491–509.

Fuchs, L., Powell, S., Seethaler, P., Cirino, P., Fletcher, J., Fuchs, D., & Zumeta, R. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology, 101,* 561–576.

Fuchs, L. S., Zumeta, R. O., Schumacher, R. F., Powell, S. R., Seethaler, P. M., Hamlett, C. L, & Fuchs, D. (2010). The effects of schema-broadening instruction on second graders' word-problem performance and their ability to represent word problems with algebraic equations: A randomized control study. *The Elementary School Journal,* 110, 440–463.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009a). *Assisting students struggling with mathematics: Response to intervention (RTI) for elementary and middle schools* (NCEE 2009–4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009b). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79,* 1202–1242.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Pyschology, 15,* 1–38.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–50). Cambridge: MIT Press.

Jayanthi, M., & Gersten, R. (2011). Effective instructional practices in mathematics for tier 2 and tier 3 instruction. In R. Gersten & R. Newman-Gonchar (Eds.), *Understanding RTI in mathematics: Proven methods and applications* (pp. 109–125). Baltimore: Paul H. Brookes. (Copyright by Paul H. Brookes).

Jitendra, A. K., & Star, J. R. (2012). An exploratory study contrasting high- and low-achieving students' percent word problem solving. *Learning and Individual Differences, 22,* 151–158.

Jitendra, A. K., Griffin, C. C., Haria, P., Leh, J., Adams, A., & Kaduvettoor, A. (2007). A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 99,* 115–127.

Jitendra, A. K., Star, J. R., Starosta, K., Leh, J. M., Sood, S., Caskie, G., & Mack, T. R. (2009). Improving seventh grade students' learning of ratio and proportion: The role of schema-based instruction. *Contemporary Educational Psychology, 34,* 250–264.

Jitendra, A. K., Star, J. R., Rodriguez, M., Lindell, M., & Someki, F. (2011). Improving students' proportional thinking using schema-based instruction. *Learning and Instruction, 21,* 731–745.

Jitendra, A. K., Dupuis, D. N., Rodriguez, M., Zaslofsky, A. F., Slater, S., Corroy, K. C., & Church, C. (2013a). A randomized controlled trial of the impact of schema-based instruction on mathematical outcomes for third grade students with mathematics difficulties. *The Elementary School Journal, 114,* 252–276.

Jitendra, A. K., Rodriguez, M., Kanive, R. G., Huang, J. P., Church, C., Corroy, K. C., & Zaslofsky, A. F. (2013b). The impact of small-group tutoring interventions on the mathematical problem solving and achievement of third grade students with mathematics difficulties. *Learning Disability Quarterly, 36,* 21–35.

Kroesbergen, E. H., & Van Luit, J. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24,* 97–114.

Milgram, R. J., & Wu, H. S. (2005). The key topics in a successful math curriculum. http://math.berkeley.edu/wu/.

National Center on Response to Intervention. (n.d.). RtI practices glossary. www.rti4success.org/index.php?option=com_content&task=view&id=634&Itemid=2.

National Council of Teachers of Mathematics. (2006). Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence. http://www.nctm.org/standards/focalpoints.aspx?id_282.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common core state standards*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.

Parker, M., & Leinhardt, G. (1995). Percent: A privileged proportion. *Review of Educational Research, 65,* 421–481.

Schmidt, W. H., & Houang, R. T. (2007). Lack of focus in the mathematics curriculum: Symptom or cause? In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 65–44). Washington, DC: Brookings Institution Press.

Schumacher, R. F., & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *Journal of Experimental Child Psychology, 111,* 607–628.

Sood, S., & Jitendra, A. K. (2007). A comparative analysis of number sense instruction in reform-based and traditional mathematics textbooks. *Journal of Special Education, 41,* 145–157.

Swanson, H., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68,* 277–321.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Synder, P. (2012). Randomized evaluation of a supplememtal grade-wide mathematics intervention. *American Educational Research Journal, 49,* 1251–1284.

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *Journal of Special Education, 32,* 207–225.

Classroom Reading Instruction for All Students

Barbara R. Foorman and Jeanne Wanzek

The response-to-intervention (RTI) approach to learning disabilities comes from a public health model that emphasizes prevention as the best form of intervention. The model is ideal in the sense that it identifies children before they fail achievement tests at a point when intervention is affordable and demonstrably effective (Foorman et al. 2003a). The Reading First Initiative—part of the No Child Left Behind Act (2001)—was a national model for RTI in the primary grades, with demonstrable impact on decoding skills (Gamse et al. 2008) and reports of comprehension gains in several states (e.g., Foorman et al. 2010). As RTI is scaled in the era of common core state standards (CCSS), there is a need to build on the successes of Reading First—a general education initiative focused on improving classroom instruction and intervening with students who fall behind—and extend this to reading in the content areas in grades 4–12. There is a need to recommit to the importance of reading instruction being aligned with a curriculum and an evidence-based scope and sequence of knowledge and skills. The context for the authors' discussion of classroom reading instruction is crucial: Trend scores in reading from the National Assessment of Educational Progress (NAEP) from 1992 to 2011 show slight gains in grades 4 and 8 (4 and 5 points, respectively) and percentages of students performing in the below-basic level of proficiency have decreased over this time period for all ethnic groups except American Indian/Alaska Native, where 53 % scored in the below-basic range. Percentages of Hispanics and Blacks scoring below basic were not far behind in 2011: 49 and 51 %, respectively. Thus, while the nation has seen some improvement in reading, these gains are not sufficient to ensure that all students can handle grade-level expectations. Yet, with the advent of the CCSS for English language arts (ELA), expectations for college and career readiness are greater than ever (National Governors Association Center for Best Practices and Council of Chief State School Officers 2012).

In this chapter, the discussion of classroom reading instruction—tier 1 in the RTI model—is organized around topics of learning to read in the primary grades and reading to learn in the content areas. But, first a brief overview of the ELA CCSS is provided with their likely effect on classroom reading instruction.

B. R. Foorman (✉) · J. Wanzek
Florida Center for Reading Research, Florida State University, 2010 Levy Ave., Ste. 100, Tallahassee, FL 32310, USA
e-mail: bfoorman@fcrr.org

J. Wanzek
e-mail: jwanzek@fcrr.org

English Language Arts Common Core State Standards

How will implementation of the ELA CCSS affect classroom reading instruction? The first thing to note is that the full name of the standards is

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_14

English language arts and literacy in history/social studies, science, and technical subjects. Thus, instead of RLA for reading language arts, ELA is for English language arts, and instead of reading in the content areas, there is literacy in the content areas. What happened to reading? Reading is a strand in the K–5 standards, along with writing, speaking, listening, and language. Literacy in the content areas is part of the 6–12 standards. Foundational reading skills are a subcomponent of the reading strand and, therefore, can easily be overlooked. In fact, foundational skills are not part of the ten anchor standards for reading. These ten anchor standards are grouped into four categories. The first category—key ideas and details—requires students to "read closely to determine what the text says explicitly and to make logical inferences from it" (ELA CCSS, p. 10). The second category—craft and structure—requires students to interpret words and phrases as they are used in the text, analyze the structure of texts, and assess point of view. The third category—integration of knowledge and ideas—requires that students evaluate arguments and claims in a text and compare and contrast themes or approaches across texts. The fourth category—range of reading and level of text complexity—requires that students "read and comprehend complex literary and informational texts independently and proficiently" (ELA CCSS, p. 10).

Thus, the anchor standards for reading in the ELA CCSS delineate the knowledge needed to read challenging (i.e., "complex") text "closely" so that the academic language and author's "point of view," "argument and specific claims" can be unpacked and compared to those in other challenging texts. A problem for teachers is that (a) this knowledge assumes students have the skills to read grade-level text and that (b) teachers understand the meaning and instructional implications of such constructs as academic language and text complexity. A brief discussion of these constructs and their instructional implications follows:

Academic Language Academic language refers to the language used in schools and other academic settings to discuss and read about literature, mathematics, science, and history/social studies. Academic language requires competence in the cognitive aspects of language as well as in the traditional components of language structure—phonology, semantics, morphology, syntax, and language use (i.e., pragmatics). The *phonological* aspects of academic language require instruction into how the writing system relates to the spoken language: (a) how the sound segments of speech (i.e., phonemes in English) map to graphic shapes (i.e., letters in English) and (b) what the appropriate stress and intonation are for English and foreign words. The *semantic* aspects of academic language (commonly called, academic vocabulary) refer to the structure of words and to their meanings—their connotations and denotations, multiple meanings across contexts, and relatedness to other words. Word structure, or *morphology,* has to do with prefixes (e.g., *pre-*), grammatical morphemes (e.g., *-ing*), and derivational suffixes (e.g., *-tion*; *-ology*). Instruction around these morphological units is required by the ELA CCSS language strand in grades 2–5. Instruction in academic vocabulary is part of the language strand at all grades, with an emphasis on "grade-appropriate general academic and domain-specific words and phrases" (ELA CCSS, p. 29–55). Examples from the ELA CCSS are provided in grades 4–5 but not in grades 6–12: (1) at grade 4, words that signal precise emotions, actions, and state of being (e.g., *quizzed, whined,* and *stammered*) or words that are basic to a particular topic (e.g., *wildlife, conservation*) and (2) at grade 5, words that signal contrast, addition, or other logical relations (e.g., *however, although, nevertheless, similarly, moreover, in addition*). Academic vocabulary relevant to the secondary grades would include content-specific words (e.g., *referendum*; *valence*) as well as words important to the scientific method, such as *describe, analyze,* and *hypothesize*.

The *syntactic* or grammatical aspects of academic language refer to the noun system, verb tenses, and complex clauses. These aspects are addressed in the ELA CCSS language strand under Conventions of Standard English. The *pragmatic* aspects of academic language refer to the use of language in context, and it is this level of language that is most relevant to the ELA CCSS speaking and listening standards. Speakers must be sensitive to what their listeners do and do not know about a topic, how to initiate conversations, take turns, maintain and change topics, and provide the appropriate amount of information in a clear way. Thus, pragmatics also includes rules of conversation or discourse, which vary depending on the context. The most frequent kinds of discourses children encounter are conversational, classroom, narrative, and event discourses. The most relevant linguistic devices for the discourse of academic language are words and phrases for conveying meaning in oral or written language, such as introductory or ending phrases, words that signal the logical relations between sentences (e.g., therefore), and cohesive devices. Common cohesive devices are: (1) overlap of noun, verb, adjective, or major ideas (e.g., "My daughter lives in New York City. She likes going to Broadway shows."); (2) use of causal verbs or particles (e.g., "He dropped the glass and, therefore, had to pick up the pieces."); and connectives (e.g., *when, because, if–then, in addition, nonetheless*).

Text Complexity The goal of the CCSS is to build college and career readiness. To do that in ELA, the standards state that students must "read widely and deeply from among a broad range of high-quality, increasingly challenging literary and informational texts" (ELA CCSS, p. 10). Exemplar texts in grades K–5 that build increasingly deeper knowledge about the human body and examples of literary and informational texts for grades 6–12 are given in the standards. Also, texts with examples of embedded performance assessments are provided in Appendix B of the standards document. To explain the definition of text complexity in anchor standard 10 (that students "read and comprehend complex literary and informational texts independently and proficiently"), the ELA CCSS describe three factors that measure text complexity (pp. 31 and 57). The first dimension is matching reader to text and task, and it includes reader variables (e.g., motivation, knowledge, and experiences) and task variables (e.g., purpose and the complexity of the task and questions asked). The second dimension is qualitative elements of text complexity: levels of meaning, structure, language conventionality and clarity, and knowledge demands. Each of these elements is rated dichotomously. For example, under knowledge demands for content/discipline knowledge, the following dichotomy is given: "Everyday knowledge and familiarity with genre conventions required" *or* "Extensive, perhaps specialized discipline-specific content knowledge required." A popular magazine would tap the everyday knowledge mentioned at the former end of this dichotomy, whereas a statistical textbook would tap the specialized discipline-specific content required at the latter end.

It can be difficult to achieve high inter-rater reliability when using such qualitative rubrics. Supplemental information for Appendix A of the ELA CCSS recognizes this fact by suggesting that two quantitative measures be used to locate a text within a grade band "because quantitative measures evaluate dimensions of text complexity—such as word frequency, sentence length, and text cohesion (to name just three)—that are difficult for a human reader to evaluate when examining a text" (National Governors Association & CCSSO 2012, p. 7). Qualitative measures are recommended for locating a text in a specific grade band once the quantitative measures have been used.

The third factor for measuring text complexity is quantitative metrics such as readability measures and the natural language dimensions of Coh-Metrix (i.e., narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion). These measures are discussed in Appendix A of the ELA CCSS, with a supplemental document detailing new research on text complexity (National Governors Association and CCSSO 2012; Nelson et al. 2012). There are six readability measures that yield a single score for text complexity. The first four are based on word difficulty and sentence length: Advantage-TASA Open Standard (ATOS) by Renaissance Learning; Degrees of Reading Power® (DRP®) by Questar Assessment, Inc.; Flesch–Kincaid in the public domain; and The Lexile® Framework for Reading by MetaMetrics. The last two measures involve computational or natural language processing techniques. The first of these is the Reading Maturity measure by Pearson Education and it utilizes (a) latent semantic analysis (LSA) to estimate knowledge of word, sentence, and paragraph meanings, and (b) other computational linguistic variables such as perplexity, sentence length, and semantic coherence. The second is TextEvaluator by the Educational Testing Service, and it computes eight dimensions of text variation, separately by genre (literary, informational, and mixed): syntactic complexity, vocabulary difficulty, level of abstractness, referential cohesion, connective cohesion, degree of academic orientation, degree of narrative orientation, and paragraph structure. In the Nelson et al. (2012) study, all of the six readability measures were reliably and, generally, highly correlated with grade-level and student performance-based measures of text difficulty across a range of text sets and reference measures. Also, all were comparable in their prediction of student outcomes and all showed increases in text complexity through high school to college and career readiness.

Educational Implications of the ELA CCSS The ELA CCSS are significantly different from existing state standards for reading/language arts. Porter et al. (2011) employed a recognized content analysis procedure, the surveys of enacted curriculum (SEC), to compare the CCSS with state content standards. The SEC defines content as the intersection of topics and cognitive demand. Porter et al. found that the ELA CCSS require teachers to place less emphasis on memorization and a much greater emphasis on analysis. This emphasis on analysis is apparent in the instructional priorities the ELA CCSS place on:

- Building knowledge through content-rich nonfiction and informational texts.
- Reading and writing grounded in evidence from text.
- Regular practice with complex texts and its academic vocabulary.

Even struggling readers are expected to have opportunities to encounter grade-level complex text. But, fortunately, there is recognition that such students will need supplemental materials and extra assistance with foundational reading skills. The next section discusses research-based practices associated with learning to read in the primary grades.

Classroom Reading Instruction in the Primary Grades

What is abundantly clear from research (e.g., Foorman et al. 1998; Mathes et al. 2005; Simmons et al. 2008) and consensus documents (National Research Council 1998; National Institute of Child Health and Human Development 2000) is that explicit instruction in the alphabetic principle of how letters map to sounds in English is necessary to learn to decode and prevent reading difficulties. However, mastery of the alphabetic principle must be coupled with construction of meaning—at the word, sentence, and text level—if comprehension is to occur (Foorman and Connor 2011; Rayner et al. 2001). Thus, reading/language arts instruction in the primary grades must consist of (a) mastering the encoding and decoding of the alphabet and (b) building oral and academic language and written language comprehension (American Educational Research Association 2009). This section starts the discussion of primary-grade reading instruction with oral and academic language skills because of their centrality to the comprehension of written language (i.e., reading comprehension). Then,instruction in decoding and encoding, word analysis, and text reading are discussed.

Teaching Academic Language Skills Children learn oral language skills incidentally at home from caregivers from the moment of birth. Yet, the longitudinal research of home language by Hart and Risley (1995) shows that by age 4, children growing up in families living in poverty have half the vocabulary and a quarter of the adult language directed to them as in middle-class homes. Also, the language that children living in high-poverty homes are exposed to and produce is *not* academic language; it is *not* the language of books that middle-class children acquire from their print-rich environment (Hart and Risley 1999). As a consequence, it is of paramount importance that teachers in prekindergarten and primary grades devote considerable time and attention to explicit teaching of academic language and in engaging children in discussions that use academic language. Specifically, teachers need to teach *inferential language, narrative language,* and *word knowledge*.

Inferential language is the language of reasoning, of forming hypotheses, predictions, conclusions, or judgments about oral or written language. Reading aloud informational or narrative text is an excellent way to build skill in inferential language if teachers ask open-ended questions that challenge students to grapple with the author's intent and the ideas in the text and to formulate statements reflective of such things as causal connections, narrative structure, and the meaning of figurative language (e.g., Coyne et al. 2004a; Lever and Sénéchal 2011). For example, after reading a passage about salmon, the teacher might ask why salmon swim upstream at certain times in the year. After reading a story about a pioneer family that traveled west in a Conestoga wagon, the teacher might ask the students to problem solve about finding food and shelter during the journey or to explain how the pioneer family's life is different from their own. Throughout such discussions, teachers need to encourage students to elaborate on their responses in complete sentences. Teachers are more consistent and effective in *not* missing opportunities to build inferencing skills when they employ scripts to guide and scaffold discussions (e.g., Foorman et al. 2003b; Justice et al. 2005; Simmons et al. 2007).

Narrative language is the language used to produce or comprehend literary text. It refers to the language used to discuss the structure of narration as well as its internal linguistic elements. The structural elements are such things as the characters, setting, goal, conflict/problem, plot/action, and resolution. The internal linguistic elements are the connectives, noun phrases, verb phrases, and pronoun references discussed above in the section on the discourse of academic language in the ELA CCSS. After reading a story, the teacher should ask students to summarize the main ideas, actions, and important details. Teachers should model a good summary (e.g., includes a beginning, middle, and an end) and scaffold discussion of how particular words aid story understanding (e.g., "because" signals that one event caused another). Additionally, teachers should prompt students if they omit important details during retelling (e.g., the family's motivation for traveling west) and point out when they provide irrelevant details (e.g., perhaps the fact that the family traveled in a Conestoga wagon; Coyne et al. 2004a; Lever and Sénéchal 2011; Simmons et al. 2007). Helping students to attend to relevant details and filter out irrelevant details is central to successful performance in all content areas.

Word Knowledge or Academic Vocabulary As stated above, students need to be explicitly taught academic vocabulary central to the meaning of text. Locating the words in text, providing definitions, and then extending the words' meaning through extension activities is a successful instructional strategy (Apthorp 2006; Coyne et al. 2009; Lever and Sénéchal 2011; Simmons et al. 2007) but not necessarily one that has lasting impacts on vocabulary breadth or reading comprehension (e.g., Apthorp et al. 2012; Goodson et al. 2011). One possible means of growing vocabulary size and impacting reading comprehension is to teach about the structure of words—their morphology (e.g., Foorman et al. 2012). This potential for morphemic instruction is addressed in the next section, but first teaching alphabetics is discussed.

Alphabetics English has an alphabetic orthography that receives bad press for its opacity, that is, for the lack of transparency between symbol and sound. However, 69 % of monosyllabic English words—those Anglo-Saxon words most used in beginning reading instruction—are consistent in their letter to pronunciation mapping (Ziegler et al. 1997) and, therefore, relations between letters and their pronunciation should be taught directly, with opportunity to practice identifying and pronouncing sounds in words in connected text. A first step in teaching kindergarteners to read is to help them become aware of the segments of sound in speech and how to manipulate them. This phonological awareness starts with larger units within words, such as syllables and moves to blending and segmenting the smallest meaningful unit of sound, phonemes. Elkonin boxes can be used to link phonemic awareness with letter–sound instruction by first having students put tiles in sequential boxes to represent sounds (e.g., X-X-X for "mat") and then having students pull alphabetic letters into the boxes to encode (i.e., spell) the word. This process of segmenting sounds in speech and mapping them to the appropriate letter is the first step toward learning the sound-spelling patterns of English that a good phonics program organizes into a sequential sequence called a scope and sequence. An effective phonics program will explicitly teach students the following: (1) to decode and write common sound-spelling patterns, (2) to blend sound-spelling patterns from left to right within a word to produce a recognizable pronunciation, (3) to practice decoding words with the sound-spelling patterns in isolation and in text, (4) to recognize high-frequency irregular words holistically (e.g., *of, to, was*), and (5) to recognize a few nondecodable words that are essential to the meaning of a text as whole words (e.g., tyrannosaurus rex). There is tremendous variability in the extent to which core reading programs provide students the opportunity to practice reading text containing words with the sound-spelling patterns taught in the phonics lesson (Foorman et al. 2004). A good list of programs with strong levels of evidence that do this well may be found on the What Works Clearinghouse (www.whatworks.ed.gov).

Phonics programs in first grade typically introduce short- and long-vowel sounds, consonant digraphs (e.g., *sh, ch, ng*), blends (*cl, str, -nd*), and silent consonants (e.g., *wr*, *kn*, *-mb*). An important insight is for students to learn that they can place more than one letter in the Elkonin boxes described above because more than one letter can represent a single phoneme. Explicit instruction in "*r*-controlled" vowels (e.g., *ir* in *bird*), diphthongs (e.g., *oi* in *oil* and *oy* in *boy*), and vowel-consonant-*e* (e.g., *pave, grave*) should be provided. An effective phonics program will post sound-spelling cards with frequent spellings for the vowel sounds. For example, long vowels have many spellings: *bake, main, clay, baby, eight, great, vein, they* for long *a*; *dime, pie, light, rifle, by, buy,* and *heist* for long *i*. One pedagogical strategy for handling the vowel inconsistency in English orthography is to establish a "set for diversity" (Gibson and Levin 1975). For example, the diphthong ow/*ow* sounds the same in the words *cow, out,* and *bough*. Students can make a "word wall" of words that represent this diphthong. Similarly, there are many spelling patterns for the vowel sound in *moon,* which students could illustrate on a word wall (e.g., *tube, suit, blue, chew, ruby*).

Word Analysis An advantage of taking a sound-spelling approach to teaching reading is that the rime pattern, which is the medial vowel and remaining consonants, is a stable orthographic unit in English that can anchor the pronunciation of medial vowels. For example, if a person hears the word "shelf" and is asked to write it, only one rime pattern would come to mind: *-elf*. However, if the person hears "sneer" and is asked to write it, there are four relevant rime patterns: *-eer* (*sneer, deer*); *-ear* (*hear, near*); *-ier* (*tier, pier*); and *–ere* (*here, mere*). Again, having students demonstrate their understanding of such sound-spelling patterns by creating word walls will facilitate their word recognition as they read and, of course, aid in their spelling accuracy (Foorman et al. 2003a).

Few core reading programs provide instruction in dividing multisyllabic words, spelling rules and conventions, or morphology—all of which are contained in effective spelling programs. However, few districts or schools purchase separate spelling programs (Foorman et al. 2004). Lack of explicit, systematic instruction in these areas is tragic because it deprives students of an arsenal of word analysis strategies for recognizing multisyllabic words, which may have a long-term negative effect on comprehension as text complexity increases. After all, English is primarily morphophonemic (Chomsky and Halle 1968), meaning that the spelling of the base word is preserved and pronunciation is shifted. Examples are *sign* in *signal* and *vine* in *vineyard*. Thus, instruction in closed, open, and stable syllable types is recommended, as well as in the three syllable types already discussed (i.e., vowel–consonant-*e*; vowel team/diphthongs, *r*-controlled). Closed syllables are consonant-vowel-consonant (CVC) patterns (e.g., *milk, fan-tas-tic*) and account for 43 % of all syllables (Stanback 1992). Open syllables are consonant-vowel (CV) patterns (e.g., *she, si-lent*) and account for 29 % of all syllables (Stanback 1992).

Core reading programs typically include instruction in a few frequent contractions (e.g., *I'll, won't, she's*), prefixes (e.g., *un*-; *ex*-), and suffixes (e.g, *-ed, -ing*). What is missing, however, is a systematic presentation of word structure (i.e., morphology) beginning in second and third grades and continuing through the elementary grades. Such a presentation would provide sequential instruction on grammatical morphemes that underlie the verb tense system in English and include instruction on related spelling conventions at the same time (e.g., changing *y* to *i* in *try* to *tried*; doubling final consonants before adding *–ing,* as in *hitting*). Such a presentation would also include systematic instruction on derivational morphemes, the meaningful units which change the meaning of the base word. Examples of derivations are the Latin suffix *–ion* (*act* to *action*) and the Greek suffix *–ology* (*meteor* to *meteorology*). Carlisle and Stone (2005) point out that words that contain no shifts between base and derived words are easier to read (e.g., *suit* to *suitable*), in contrast to words with one shift in phonology (e.g., *satire* to *satirical*; *caliber* to *calibration*) or one shift in orthography (*pity* to *piteous*; *cry* to *cries*; *begin* to *beginning*; *proceed* to *procedure*; *secure* to *security*). Not surprising, the most difficult words to read are those with shifts in both phonology and orthography: two-shift words such as *mature* to *maturation*; *theory* to *theoretical* (Carlisle and Stone 2005). Hence, explicit instruction in morphology is one way to facilitate comprehension as text complexity increases.

Awareness of morphemic structure is associated with improvements in reading comprehension. Nunes et al. (2012) found that morphological awareness measured when students were ages 8 and 9 predicted their reading comprehension at ages 12 and 13. In a study with 4780 students in grades 3–10, Foorman et al. (2012) found that morphological knowledge added 2–9 % unique variance beyond prior reading comprehension in predicting spring reading comprehension at the student level and that over 90 % of the large variability between classrooms was explained by performance on component skills of morphological knowledge, text reading efficiency, and spelling.

Text Reading As students are learning to read, they need to read a variety of connected text every day with and without feedback in order to build accuracy, fluency, and comprehension (e.g., Coyne et al. 2004b; Denton et al. 2010; Scanlon et al. 2010). The particular texts selected, the format for grouping students (e.g., with the whole class together, in small groups, in pairs, or individually), and the instructional strategies selected will depend on the instructional goal. For example, in order to build inferential language skills and academic vocabulary about the solar system, a second-grade teacher might read a science text to the whole class and engage the students in group discussion and follow-up research activities in small groups. In order to practice decoding skills, a first-grade teacher might listen to students read aloud in a small group from instructional-level text, pausing to provide corrective feedback and have students write words on their dry-erase or white boards. Instructional-level text is a text in which a student does not make more than 10 % word-reading errors so that the teacher can scaffold the correct pronunciation of the word, and the student does not become frustrated (O'Connor et al. 2010).

Partner reading with a more capable peer is also a good format to meet the goal of improving word identification or building fluency in instructional-level text (e.g., O'Connor et al. 2007). Having students echo the teacher's reading can also serve these goals, as can choral reading in a group as long as all students are engaged. Independent silent reading in independent-level text is important for building reading comprehension skills, as long as understanding can be demonstrated through discussion, written summaries, or graphic organizers. Independent readers in the primary grades will also benefit from training in the text structures of informational text, such as the compare–contrast structure (Williams et al. 2009). In fact, a goal of the ELA common core standards is for primary grade students to demonstrate proficiency in comparing and synthesizing ideas across a range of informational sources. In sum, primary-grade teachers need to provide daily opportunities for their students to read connected text with attention to matching instructional objectives and strategies with appropriate selection of texts and grouping of students to achieve accuracy, fluency, and comprehension.

Classroom Reading Instruction in the Secondary Grades

It is imperative that reading instruction for all students continue beyond the primary grades. National and international data suggest there are a wide range of students who are not reading at proficient levels in the secondary grades (National Center for Education Statistics 2009; Organization for Economic Cooperation and Development 2011). There may be a number of reasons why literacy levels for adolescents are low. Some students may not have received effective reading instruction or intervention in the earliest grades and, thus, continue to struggle with even basic reading. Other students may have mastered basic foundational skills but struggle with the higher-level reading vocabulary, comprehension, and critical thinking skills that are emphasized in the upper grades. In many cases, explicit instruction and opportunity to practice reading and comprehending text fades after third grade, and many students may struggle to generalize their early literacy knowledge and skills to the wide variety and complexity of texts they are expected to read and understand independently throughout the secondary grades. Although effective instruction in the foundational skills as well as higher-level text-processing skills are an important part of elementary instruction, successful acquisition of the early literacy skills expected in kindergarten through third grade does not necessarily provide students with all the essential knowledge and skills needed to remain proficient readers in the upper grades (Snow et al. 2007).

A recent examination of students taking the American College Testing (ACT) (2006) suggested the largest difference for students who demonstrated college readiness was their ability to read and understand complex texts. These complex texts included: (a) subtle and deeply embedded relationships and interactions occurring among ideas or characters in the text; (b) rich text with a substantial amount of highly intricate or involved information being conveyed; (c) elaborate, implicit, and sometimes unconventional, organization or structure to the text; (d) tone and use of language that is sophisticated, unfamiliar, and/or multidimensional; (e) vocabulary that is demanding and highly dependent on the context; and (f) ambiguous, multilayered, or implicit author purpose (ACT 2006). As students progress through the grade levels, text length and the density of information within the text increases exponentially, requiring advanced reading practices to read and understand. In addition, students in middle and high school are expected to read text not only in ELA classes but also as a medium for learning and thinking critically about science, social studies, and mathematics. Thus, one unique challenge in the upper grades is the variety of text, contexts, disciplines, and purposes for reading and understanding for which students must be proficient (Carnegie Council on Advancing Adolescent Literacy 2010; Heller and Greenleaf 2007; Lee and Spratley 2010).

Systematic instruction in the high-level, mature-reading practices needed to understand and analyze these complex texts across content areas simply cannot fully occur in the primary grades and must continue to be addressed in the upper grades as students are progressing as readers and learners. Therefore, adolescent readers require effective tier 1 instruction that goes across the content areas and includes disciplinary literacy instruction that is embedded throughout the content (Biancarosa and Snow 2006; Kamil et al. 2008; Shanahan and Shanahan 2008; Torgesen et al. 2007). As noted earlier, it is for this reason that the ELA CCSS standards provide specific literacy standards in grades 6–12 for not only ELA but also social studies, science, and technical subjects.

High-Quality Tier 1 Reading Instruction in the Secondary Grades

Previous research provides some guidance for effective reading instructional practices at the secondary level. The Institute of Education Sciences recently provided a guidance document summarizing research evidence for classroom and instructional practices for adolescent literacy (Kamil et al. 2008). Research from Grades 4 and up was examined to develop recommendations for effective instructional practices for adolescent literacy, and the level of research evidence supporting each recommendation was detailed. In the summary, a practice was considered to have strong evidence if there were consistent findings from causal studies with generalizable samples that the practice led to better reading outcomes for students. Moderate evidence for a practice referred to strong causal *or* generalizable evidence but some ambiguity in one area. Low evidence was characterized by strong findings or theories in related areas or studies that did not meet moderate or strong evidence standards. The review identified strong evidence in the research for two, tier 1 literacy practices: (a) explicit vocabulary instruction, and (b) explicit comprehension strategy instruction. In addition, moderate evidence was noted in the research for instruction that provided opportunities for extended discussion of text and increased student motivation and engagement in literacy learning (Kamil et al. 2008). Similarly, the Center on Instruction has provided a guidance document for secondary content area teachers using key research studies (Torgesen et al. 2007). Torgesen and colleagues recommended tier 1 instruction include: (a) high standards for text, conversation, questions, and vocabulary, (b) increased amounts of explicit instruction and support in the use of effective comprehension strategies throughout content classes, (c) increased amount and quality of extended discussion of reading content, and (d) use of a variety of practices to improve motivation and engagement with reading (Torgesen et al. 2007).

Explicit Vocabulary Instruction Vocabulary knowledge includes receptive and expressive abilities to understand and use a variety of words in listening, reading, speaking, and writing activities in order to communicate and comprehend effectively. Vocabulary knowledge predicts differences in reading ability even for adults (Swanson and Hsieh 2009). In addition, the development of student vocabulary is related to student reading comprehension of challenging text (Cunningham and Stanovich 1997; National Reading Panel 2000; Stahl and Fairbanks 2006). As text complexity increases through the grades so too does the number of unique words and concepts students must accurately understand including content-area-specific, academic language (Harmon et al. 2005). Therefore, direct, explicit, vocabulary instruction with an intentional focus on developing students' knowledge of word meanings and academic language is one important means for improving student vocabulary knowledge and use through the grades (Biemiller 2001; Blachowicz et al. 2006). This explicit instruction of targeted words can also serve to improve students' access and understanding of more complex text.

The research on vocabulary development suggests that increasing the amount of explicit vocabulary instruction results in more learning of word meanings by students (Beck and McKeown 2007; Biemiller and Boote 2006; Stahl and Fairbanks 2006) and can also serve to improve comprehension of text (National Reading Panel 2000). In a meta-analysis of the effects of vocabulary instruction, a mean effect size of 0.97 was noted for reading comprehension (Stahl and Fairbanks 2006). In addition, a correlation of $r=0.65$ was reported between time allocated to vocabulary instruction and student reading comprehension outcomes. However, Elleman et al. (2009) noted these large effects were found on researcher-developed comprehension measures which had a high degree of overlap with the words targeted during intervention, whereas broader, standardized comprehension measures did not show statistically significant effects.

The following key areas of instruction have been identified in the research on effective vocabulary instruction for the upper grades (Baker et al. 1998; Beck et al. 1987; Nagy and Scott 2000; Stahl and Fairbanks 2006):

- Teaching students the meaning of new words that students will encounter in text in a targeted and explicit fashion.
- Providing definitional and contextual information when teaching word meanings.
- Providing multiple exposures and practice with new words in a variety of contexts.
- Teaching students independent strategies for determining the meanings of additional words not explicitly taught.

In selecting target words for explicit vocabulary instruction, it is important to consider the words and concepts that are essential to understanding and learning the content and that occur frequently in the content and/or provide overarching conceptual information to build content knowledge. Rather than learning definitions of the target words through rote memory, instructional activities need to facilitate a thorough, conceptual understanding of the meanings of words in context. When students understand how the new words are used in oral and written language, they can better comprehend text. Providing multiple, in-context practice opportunities through text and content discussions provides students with the varied practice needed to ensure that the new words become a part of their vocabulary and comprehension. As students improve their knowledge of vocabulary in various content areas, teachers can better relate new concepts to these previously taught contexts and concepts, deepening student understanding of the vocabulary as well as content (Twyman et al. 2006).

It is not possible to target and teach students all the meanings of words in the English language. Thus, instruction in the application of independent strategies for determining and learning meanings of new or unfamiliar words is also vital. There is evidence to suggest that explicit instruction in using dictionaries, morphemic analysis, and the use of context to determine word meanings can assist students in gaining knowledge of unfamiliar words independently (National Reading Panel 2000) and should accompany (not replace) targeted, explicit word meaning instruction.

Explicit Comprehension Instruction Constructing meaning from complex text across a variety of content areas requires proficient and sophisticated application of a number of processes (Snow 2002). Comprehension instruction can be one of the more challenging reading components to address because much of what is needed to understand text happens through a complex system of thought processes that are generally not transparent to a student developing as a reader. Thus, it is necessary to make these thought processes for understanding text explicit for students during instruction. Whether teachers are introducing a comprehension skill or a more complex comprehension strategy, ensuring that instruction is provided prior to student practice or assessment is the key to effective and efficient student learning.

Comprehension instruction requires explicitly teaching students the purpose and process for effectively using text reading practices that will help improve their understanding of text (Almasi 2003; Block 1993; Dole et al. 1996; Klingner et al. 1998a; Lysynchuk et al. 1990; Perfetti et al. 2005). Explicit, effective comprehension instruction includes: (a) setting a purpose for use of the comprehension skill or strategy, (b) teacher modeling of the skill or strategy including how to apply the strategy and when to use it successfully, (c) extensive, in-context practice opportunities to use the skills and strategies and receive feedback, and (d) faded teacher support to transfer the learned skills and strategies to independent student use. Ultimately, student independence in using the comprehension practices during reading is the goal, but when independent practice begins too soon, students may not be able to successfully apply the skill or strategy or may be unable to master and retain the use of the strategy when reading. Comprehension instruction is thus an active process between the teacher and students that cannot be accomplished by simply assigning reading to students and checking students' comprehension through questions or assessments. Instead, teachers must provide instruction in how to accomplish the comprehension practice, have students demonstrate their use of the practice, and provide prompting and corrective feedback as needed to ensure effective use.

There is a considerable amount of research examining comprehension instruction that can improve reading outcomes for students with and without reading difficulties (e.g., Alfassi 2004; Armbruster et al. 1991; Block 1993; Dole et al. 1996; Edmonds et al. 2009; Graves et al. 1983; Kim et al. 2004; Klingner et al. 1998a; Mastropieri et al. 1996; Palinscar and Brown 1984; Pressley 1998; Swanson 1999). Specifically, these research studies indicate reading achievement is increased at all grade and ability levels when students are taught to: (a) activate and build background knowledge; (b) preview, predict, and confirm predictions; (c) retell and summarize text; (d) ask questions about the reading; (e) use graphic organizers to arrange, categorize, and/or relate key information in text; (f) engage in inferential reasoning; (g) visualize; and/or (h) monitor and repair comprehension while reading. As a result, teaching students how to use these practices when reading is an important part of tier 1 instruction in the upper grades.

Effective comprehension instruction includes application of some skills and strategies that are applicable across many texts or content areas such as the use of graphic organizers, summarizing key ideas, and generating questions about the text to monitor understanding of the reading. Additionally, there may be some comprehension practices that are specific to a particular content area or the content area-specific text including the specific prior knowledge required, relating of ideas to predict or generate content area-specific questions, or the appropriate use of supportive evidence. For example, identifying evidence to support a point or main idea may differ from one content area to another. The types of evidence used to support or identify a main idea in a literary piece may be considered inadequate for supporting evidence of the main points in a historical document in a world history class. In addition, text complexity and text relation to student background knowledge are ever changing within and across content areas, making comprehension a dynamic process. Thus, students need instruction, practice, and feedback applying comprehension skills and strategies across a variety of texts that may differ in content, explicitness of information presented, ambiguity of main ideas, distance between relevant ideas, and density of key ideas.

The greatest effects for improving comprehension may come from teaching multiple strategies across content areas (Alfassi 2004; National Reading Panel 2000; Taylor et al. 2005). However, students must understand not only how to apply each strategy but why and when to apply the strategy; thus providing them with a range of tools that they can adapt and employ in appropriate situations to access and understand a variety of complex text across content areas. Therefore, providing supported reading during the content areas to teach reading routines where students consider or build relevant background knowledge prior to reading, stop regularly to question and judge understanding of the text during reading, and synthesize text information and provide evidence for answers, inferences, and opinions after reading can build student proficiency for comprehending complex text.

Extended Discussion of Text and Increasing Engagement In addition to the strong research evidence for implementing explicit vocabulary and comprehension instruction at the secondary levels, Kamil et al. (2008) noted moderate evidence for providing opportunities for extended discussion of text as well as increasing student motivation and engagement in literacy learning. The moderate evidence in the research for these practices suggests student outcomes may be enhanced when opportunities are provided for extended discussion of text and student motivation and engagement for literacy learning are integrated in tier 1 instruction. Both discussion and motivation techniques lend themselves easily to integration with explicit vocabulary and comprehension instruction across the content areas.

High-quality discussions can provide an avenue for students to think critically about text, engage in reading, and deepen comprehension and learning of the content and text (Applebee et al. 2003; Michaels et al. 2002; Murphy et al. 2009). Discussions that are teacher directed offer the opportunity to support students in making key connections and inferences between the text and prior knowledge, and forming sophisticated representations of the content and text that can increase text understanding as well as retention of information. Text discussion is often integrated with vocabulary or comprehension strategy practices (Klingner et al. 1998b; Palinscar and Brown 1984). Kamil et al. (2008) recommend the need for engaging materials and stimulating questions as well as a structured routine for holding the discussion in order to foster the type of discussion that is beneficial to student comprehension.

Increasing motivation for students to read and engage in text may also be necessary in order for students to participate actively in the instruction and practice related to vocabulary and comprehension and to reap the benefits of these practices (Guthrie et al. 2004; Snow 2002; Wigfield et al. 2008). Building student motivation, including goal setting, self-directed learning, collaboration opportunities, and strategy instruction, can lead to increased engagement (Guthrie et al. 2000). In addition, practices such as including choices and autonomy in identifying text or assignments are recommended practices to embed in the key instructional practices addressed earlier in this chapter (Guthrie et al. 1999).

Current Tier 1 Practices in the Secondary Grades

Historically, studies examining instruction through the content areas for students in the secondary grades have noted a lack of the evidence-based instruction in reading, vocabulary building, and comprehension strategies needed to assist students in engaging in and understanding complex text (Durkin 1978–1979; Pressley 2004; Scott et al. 2003). With the recent renewed focus on adolescent academic literacy instruction (Biancarosa and Snow 2006; Carnegie Council on Advancing Adolescent Literacy 2010; Heller and Greenleaf 2007; Lee and Spratley 2010), Swanson et al. (in progress) recently examined current practices in two content areas, ELA and social studies in grades 7–12 to document the use of text in instruction as well as the strategies teachers provided for reading and understanding text. The literacy practices of 12 experienced social studies teachers and 9 experienced ELA teachers in grades 7–12 were documented across three districts in two states. A total of 79 social studies class periods and 58 ELA class periods were observed. The occurrence of vocabulary instruction or practice and comprehension instruction or practice was documented for each class period. In addition, the amount of time students or teachers spent reading text was recorded.

In the area of vocabulary, providing students with definitions for unfamiliar words was the most common practice in both the social studies and ELA classes, occurring in 52 % of the social studies observations and 60 % of the ELA observations. Small amounts of morphology and context clue work were noted during social studies (4–8 % of observations) and ELA (12–19 % of observations). Comprehension instruction was noted in both social studies and ELA classes. Teacher questioning was the most common comprehension practice in both content areas, noted in approximately 33 % of the social studies observations and about 50 % of the ELA observations. The practices of activating or building background knowledge, previewing text, reading comprehension strategy instruction, and extended discussion were also noted to a lesser degree in some of the observations in both social studies and ELA (Swanson et al. in press).

Despite these practices, Swanson et al. (in press) reported text reading occurred a limited amount of time in both social studies and ELA. Text reading in social studies classes occurred approximately 10 % of the total time observed (3925 min of observation), while text was read approximately 15 % of the total observed time in ELA (3283 min of observed time). Not surprisingly, the large majority of text in social studies was expository, while the majority of text in ELA

Table 1 Implications for tier 1 instruction

Primary grade tier 1 reading instruction	
Recommendation	Key areas of instruction
Teach academic language skills	Teach inferential language, narrative language, and word knowledge
Provide explicit instruction in the alphabetic principle	Teach phonological awareness; letter names and sounds and blending to produce a recognizable pronunciation; decoding and encoding of common sound-spelling patterns; decoding in isolation and in text; recognition of high frequency, irregular words
Teach word analysis skills	Teach syllable patterns; word structure (prefixes; grammatical and derivational morphemes)
Text reading	Ensure daily reading of a variety of text with and without feedback to build accuracy, fluency, and comprehension
Secondary grade/content area tier 1 reading instruction	
Recommendation	Key areas of instruction
Provide explicit vocabulary instruction	Target general and domain-specific words students will encounter in text, are essential to learning the content, occur frequently in the content, and/or provide conceptual information
	Provide definitional and contextual information to teach word meanings
	Provide multiple practice opportunities for engaging with new words using a variety of contexts
	Teach independent strategies for determining the meanings of unknown words not targeted in instruction
Provide explicit comprehension instruction	Teach students active processes for understanding text including summarizing, generating questions, organizing information in graphic organizers, inferential reasoning, visualizing, and monitoring comprehension
	Set a purpose for the use of specific comprehension practices
	Model comprehension practices for students including how to apply the practice and when to use each practice
	Provide extensive, meaningful, practice opportunities with feedback for applying the new comprehension practice in reading
	Organize comprehension practices into a reading routine for students to apply when reading in a content area
Provide opportunities for extended discussion of text	Engage students in teacher-directed discussions of text that require critical thinking and have an explicit goal
	Ensure discussion stays focused on academic content and objectives for the lesson/text
	Encourage students to explain their thinking and support their ideas with text evidence
Increase student motivation and engagement in literacy learning	Build student interest for literacy learning; set a meaningful purpose for instruction; teach students to set goals for learning; provide opportunities for self-directed learning or student choice when appropriate; provide opportunities for peer collaboration

was narrative though some expository was also utilized. Social studies teachers typically read the text to students or had students engage in small-group or partner reading. ELA teachers most commonly read the text to students or played an audio recording of the text. Independent reading was also observed to a lesser degree in both content areas. Teachers in both content areas typically checked on student comprehension and understanding of the text by having students answer oral questions (Swanson et al. in press).

Given that students' experience with complex and conceptually demanding text in a variety of areas is a strong predictor of success in early college courses (ACT 2006), it is striking how little text was read and discussed in these content area classes at the secondary level. It is particularly concerning that often when text was used in instruction, the teacher (or an audio recording) read the text to the students. Implementation of CCSS will require increases in access to and instruction in reading and comprehending challenging text for all students as a part of effective tier 1 instruction in the upper grades. It is imperative that tier 1 instruction for older students provide a focus on reading, vocabulary and academic language, background knowledge, making inferences, and comprehension strategy use as these practices best predict students' comprehension of text (Cromley and Azevedo 2007). Current adolescent reading achievement data suggest students in the secondary grades are not presently proficient, independent readers of complex text. As a result, tier 1 instruction in the upper grades must provide advanced instruction in reading and understanding text and support students to apply effective practices for independent reading and comprehension in the types of text they will encounter in college and the workforce.

Conclusion

The status of tier 1 reading instruction is at a critical juncture. We, as a nation, have committed to an educational goal of college and career readiness for all students. For the vast majority of states, this goal is instantiated in the form of the ELA CCSS. The demands of these standards on students' analytic and academic language skills are high. Are we ready?

To meet this challenge an in-depth and on-going professional development for all K–12 teachers in the ELA CCSS is needed. This includes not only an emphasis on implementing the standards and the integration across strands at each grade level but also an understanding of how the standards align vertically to build knowledge across grades. Special attention must be paid to what instructional practice looks like for (a) developing academic language, (b) careful reading to ensure comprehension of complex text (i.e., what the ELA CCSS calls "close reading"), and (c) determining point of view and evaluating claims and evidence.

It is imperative that as the ELA CCSS or college and career readiness standards are implemented that the strong evidence base that has been gathered in foundational reading skills and effective strategies for improving reading comprehension in the content areas is not disregarded. The teachers know how to teach students to be good decoders and spellers, but, of course, knowing and doing are not the same, and, even when decoding and spelling are taught, they are not always taught well. There are many promising approaches for building vocabulary, improving reading comprehension, and increasing student motivation and engagement in literacy learning. Let us continue to implement these evidence-based practices and at the same time raise expectations for effective classroom reading instruction for all students. The authors conclude with a table of recommendations for tier 1 reading instruction that summarizes the practices discussed in this chapter. (see Table 1)

Acknowledgment The research reported here was supported by the Institute of Education Sciences, the US Department of Education, through sub awards to two grants: (1) a sub award to Florida State University (FSU) from Grant R305F100005 to the Educational Testing Service as part of the Reading for Understanding Initiative and (2) a sub award to FSU from Grant R305F100013 to the University of Texas at Austin. The opinions expressed are those of the authors and do not represent views of the Institute, the US Department of Education, the Educational Testing Service, The University of Texas at Austin, or The Florida State University.

References

ACT. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City: Author.

Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. *Journal of Educational Research, 97,* 171–184. doi:10.3200/JOER.97.4.171–185.

Almasi, J. F. (2003). *Teaching strategic processes in reading*. New York: Guilford.

American Educational Research Association. (Winter, 2009). Ensuring early literacy success. *Research Points, 6*(1), 1–4. http://www.aera.net/Portals/38/docs/Publications/Ensuring%20Early%20Literary%20Success.pdf.

Applebee, A., Langer, J., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Education Research Journal, 40,* 685–730. doi:10.3102/00028312040003685.

Apthorp, H. S. (2006). Effects of a supplemental vocabulary program in third-grade reading/language arts. *Journal of Educational Research, 100*(2), 67–79.

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness, 5*(2), 160–188.

Armbruster, B. B., Anderson, T. H., & Meyer, J. (1991). Improving content area reading using instructional graphics. *Reading Research Quarterly, 26*(4), 393–416.

Baker, S. K., Simmons, D. C., & Kame'enui, E. J. (1998). Vocabulary acquisition: Research bases. In D. C. Simmons & E. J. Kame'enui (Eds.), *What reading research tells us about children with diverse learning needs: Bases and basics* (pp. 183–217). Mahwah: Lawrence Erlbaum.

Beck, I. L., & McKeown, M. G. (2007). Increasing young children's oral vocabulary repertories through rich and focused instruction. *Elementary School Journal*, *107*, 251–271. doi:10.1086/511706.

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147–163). Mahwah: Lawrence Erlbaum.

Biancarosa, G., & Snow, C. (2006). Reading next: *A vision for action and research in middle and high school literacy: A report to the Carnegie Corporation of NewYork* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *The American Educator, 25*(1), 24–28.

Biemiller, A., & Boote, C. (2006). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology, 98,* 44–62. doi:10.1037/0022-0663.98.1.44.

Blachowicz, C. L. Z., Fisher, P. J. L., Ogle, D., & Watts-Taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly, 41*(4), 524–539. doi:10.1598/RRQ.41.4.5.

Block, C. C. (1993). Strategy instruction in a literature-based reading program. *Elementary School Journal, 94,* 139–151.

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40,* 428–449.

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York: Carnegie Corporation of New York.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Coyne, M. D., Simmons, D. C., Kame'Enui, E. J., & Stoolmiller, M. (2004a). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145–162.

Coyne, M. D., Kame'Enui, E. J., Simmons, D. C., & Harn, B. A. (2004b). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities, 37*(2), 90–104.

Coyne, M., McCoach, D., Loftus, S., Zipoli, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal, 110,* 1–18.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325. doi:10.1037/0022-0663.99.2.311.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934–945. doi:10.1037/0012-1649.33.6.934.

Denton, C. A., Kethley, C., Nimon, K., Kurz, T. B., Mathes, P. G., Shih, M., & Swanson, E. A. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children, 76*(4), 394–416.

Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly, 31,* 62–88. doi:10.1598/RRQ.31.1.4.

Durkin, D. (1978–1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14,* 481–253.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research, 79,* 262–300. doi:10.3102/0034654308325998.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). Instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2,* 1–44. doi:10.1080/19345740802539200.

Foorman, B. R., & Connor, C. (2011). Primary reading. In M. Kamil, P. D. Pearson, & E. Moje (Eds.), *Handbook on reading research* (Vol. 4, pp. 136–156). New York: Taylor and Francis.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37–55.

Foorman, B., Breier, J., & Fletcher, J. (2003a). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*(2–3), 613–639.

Foorman, B., Seals, L., Anthony, J., & Pollard-Durodola, S. (2003b). Vocabulary enrichment program for third and fourth grade African American students: Description, implementation, and impact. In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 419–441). Austin: Pro-Ed.

Foorman, B. R., Francis, D. J., Davidson, K., Harm, M., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies in Reading, 8*(2), 167–197.

Foorman, B. R., Petscher, Y., Lefsky, E., & Toste, J. (2010). Reading first in Florida: Five years of improvement. *Journal of Research on Literacy, 42,* 71–93.

Foorman, B., Petscher, Y., & Bishop, M. D. (2012). The incremental variance of morphological knowledge to reading comprehension in grades 3–10 beyond prior reading comprehension, spelling, and text reading efficiency. *Learning and Individual Differences, 22,* 792–798.

Gamse, B., Jacob, R., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First impact study final report (NCEE 2009-4038)*. Washington, DC: National Center for Educational Evaluation & Regional Assistance, Institute of Education Sciences. (U.S. Department of Education).

Gibson, E., & Levin, H. (1975). *The psychology of reading*. Cambridge: MIT Press.

Goodson, B., Wolf, A., Bell, S., Turner, H., & Finney, P. (2011). *Effectiveness of a program to accelerate vocabulary development in kindergarten (VOCAB): First grade follow-up impact report and exploratory analyses of kindergarten impacts (NCEE 2012-4009)*. Washington, DC: National Center for Educational Evaluation & Regional Assistance, IES. (U.S. Department of Education).

Graves, M. F., Cooke, C. L., & Laberge, M. J. (1983). Effects of previewing short stories. *Reading Research Quarterly, 18,* 262–276.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., et al. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology, 96,* 403–423. doi:10.1037/0022-0663.96.3.403.

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading, 3*(3), 231–256. doi:10.1207/s1532799xssr0303_3.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92*(2), 331–341. doi:10.1037/0022-0663.92.2.331.

Harmon, J., Hedrick, W. B., & Wood, K. D. (2005). Research on vocabulary instruction in the content areas: Implications for struggling readers. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 21,* 261–280. doi:10.1080/10573560590949377.

Hart, B., & Risley, T. (1995). *Meaningful differences*. Baltimore: Brookes Publishing Co.

Hart, B., & Risley, T. (1999). *The social world of children learning to talk*. Baltimore: Brookes Publishing Co.

Heller, R., & Greenleaf, C. L. (2007). *Literacy instruction in the content areas*: *Getting to the core of middle and high school improvement*. Washington, DC: Alliance for Excellent Education.

Justice, L., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17–32.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. (U.S. Department of Education. http://ies.ed.gov/ncee/wwc.).

Kim, A., Vaughn, S., Wanzek, J., & Wei, S. (2004). A synthesis of research on graphic organizers and their effect on reading comprehension for students with learning disabilities. *Journal of Learning Disabilities, 37,* 105–118.

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998a). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal, 99,* 3–22.

Klingner, J. K., Vaughn, S., Hughes, S. T., Schumm, J. S., & Elbaum, B. (1998b). Outcomes for students with and without learning disabilities in inclusive classrooms. *Learning Disabilities Research & Practice, 13,* 153–161.

Lee, C. D., & Spratley, A. (2010) *Reading in the disciplines: The challenges of adolescent literacy*. New York: Carnegie Corporation.

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology, 108,* 1–24.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal, 90,* 469–484.

Mastropieri, M. A., Scruggs, T. E., Bakken, J. P., & Whedon, C. (1996a). Reading comprehension: A synthesis of research in learning disabilities. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 10, Part B, pp. 201–223). Greenwich: JAI Press. doi:10.1177/0741932509355988.

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*(2), 148–182.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. (2002). *Accountable talk: Classroom conversation that works (3 CD-ROM set)*. Pittsburgh: University of Pittsburgh.

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009) Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101,* 740–764. doi:10.1037/a0015576.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary process. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Mahwah: Erlbaum.

National Center for Education Statistics. (2009). *National assessment of educational progress*. Washington, DC: U.S. Department of Education.

National Governors Association & Council of Chief State School Officers. (2012). *Supplemental information for appendix a of the common core standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington DC: National Governors Association Center for Best Practices, Council of Chief State School Officers. http://www.corestandards.org/assets/E0813_Appendix_A_New_Research_on_Text_Complexity.pdf.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: Reports of the subgroups (NIH Publication No. 00-4754)*. Washington, DC: U.S. Department of Health and Human Services.

National Reading Panel. (2000). *Report of the National Reading Panel: Reports of the subgroups*. Washington, DC: U.S. Department of Health and Human Services, National Institute of Health.

National Research Council (1998). Preventing reading difficulties in young children. In C. E. Snow, M. S. Burns, & P. Griffin (Eds.), *Committee on the prevention of reading difficulties in young children, committee on behavioral and social science and education*. Washington, DC: National Academy Press.

Nelson, J., Perfetti, C., Liben, D., Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Washington, DC: Council of Chief State School Officers. http://www.ccsso.org/Documents/2012/Measures%20ofText%20Difficulty_final.2012.pdf.

Nunes, T., Bryant, P., & Barros, R. (2012). The development of word recognition and its significance for comprehension and fluency. *Journal of Educational Psychology*. doi:10.1037/a0027412.

O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children, 74*(1), 31–46.

O'Connor, R. E., Swanson, H. L., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology, 102,* 1–19.

Organization for Economic Cooperation and Development (2011). *Education at a glance 2011: OECD indicators*. Paris: OECD Publishing. doi:10.1787/eag-2011-en.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175. doi:10.1207/s1532690xci0102_1.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 227–247). Oxford: Blackwell. doi:10.1002/9780470757642.ch13.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher*, *40*(3), 103–116.

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. New York: Guilford.

Pressley, M. (2004). The need for research on secondary literacy education. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 415–432). New York: Guilford.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31–74.

Scanlon, D., Anderson, K., & Sweeney, J. (2010). *Early interventions for reading difficulties: The interactive strategies approach*. New York: Guilford.

Scott, J. A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal, 103,* 269–286.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content area literacy. *Harvard Educational Review, 78*(1), 40–59.

Simmons, D., Kame'enui, E., Harn, B., Coyne, M., Edwards, L., & Thomas, C. (2007). The effects of instructional emphasis and specificity on early reading and vocabulary development of kindergarten children. *Journal of Learning Disabilities*, *40*, 331–347.

Simmons, D., Coyne, M., Kwok, O., McDonagh, S., Harn, B., & Kame-ennui, E. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41,* 158–173.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica: RAND.

Snow, C., Porche, M. V., Tabors, P., & Harris, S. (2007). *Is literacy enough? Pathways to academic success for adolescents*. Baltimore: Paul H. Brookes.

Stahl, S. A., & Fairbanks, M. M. (2006). The effects of vocabulary instruction. In K. A. D. Stahl & M. C. McKenna (Eds.), *Reading research at work: Foundations of effective practice* (pp. 226–261). New York: Guilford.

Stanback, M. (1992). Syllable and rime patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia, 42*(1), 196–221.

Swanson, H. L. (1999). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68,* 277–321.

Swanson, H. L., & Hsieh, C. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research, 79,* 1362–1390. doi:10.3102/0034654309350931.

Swanson, E. A., Wanzek, J., McCulley, L., Stillman-Spisak, S., Vaughn, S., Summon, D., Fogarty, M., & Hairrell, A. (in press). Literacy and text reading in middle and high school social studies and English language arts classrooms. *Reading and Writing Quarterly*.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2005). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly, 40*(1), 40–69. doi:10.1598/RRQ.40.1.3.

Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S., & Lesaux, N. (2007). *Academic literacy instruction for adolescents: A guidance document from the Center on Instruction*. Portsmouth: RMC Corporation, Center on Instruction.

Twyman, T., McCleery, J., & Tindal, G. (2006). Using concepts to frame history content. *Journal of Experimental Education, 74,* 331–349.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Lutz Klauda, S., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in Schools, 45*(5), 432–445. doi:10.1002/pits.20307.

Williams, J. P., Brooke, S., Lauer, K., Hall, K., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101*(1), 1–20.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for -OUGH and the spelling for/u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, & Computers, 29,* 600–618.

Classwide Intervention Using Peer-Assisted Learning Strategies

Kristen L. McMaster and Douglas Fuchs

Current response to intervention (RTI) models specify multi-tiered approaches to prevention and intervention that begin with high-quality core instruction in the general education classroom. A majority of children are expected to thrive academically within this core instruction, with relatively small numbers of children requiring more intensive levels of intervention. In the absence of high-quality core instruction, proportions of children requiring additional support may rise considerably, straining school resources and potentially making intervention delivery and management infeasible (Burns 2010). Thus, districts and schools must make high-quality core instruction a priority in their adoption and implementation of RTI.

The imperative for high-quality core instruction is emphasized in both the general and special education laws, which specify that teachers must implement *scientific, research-based instruction* to ensure that all students are progressing toward high academic standards (Elementary and Secondary Education Act 2002) and to rule out poor instruction in eligibility decisions for students being evaluated for possible learning disabilities (Individuals with Disabilities Education Act 2004). However, few (if any) comprehensive core instructional programs have been validated through rigorous research (Gersten et al. 2009). Thus, districts and schools are advised to select core curricula that align with scientifically based principles of teaching and learning (Al Otaiba et al. 2005), and to supplement those core curricula with research-based classwide programs and interventions, particularly in classrooms with high proportions of students at risk for academic failure.

Research-based classwide interventions can help teachers ensure that students have sufficient opportunities to practice critical academic skills addressed in the core curriculum, as well as to differentiate instruction to meet a wide range of student needs (Gersten et al. 2009). To select appropriate classwide interventions, educators should consider the following questions: (a) Does the intervention *address critical academic skills* that align with important academic standards? (b) Does the intervention *provide opportunities to differentiate instruction* for low-, average-, and high-performing readers? (c) Does the intervention *have evidence of efficacy* for students like those in the classroom for which it is to be used, including students from diverse cultural and linguistic backgrounds, and students with disabilities? (d) What types of *professional development and support* are needed to ensure fidelity of implementation? (e) Is there evidence that the in-

K. L. McMaster (✉)
University of Minnesota, Minneapolis, USA
e-mail: mcmas004@umn.edu

D. Fuchs
Peabody College of Vanderbilt University,
Nashville, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_15

tervention can be *adapted to meet classroom and student needs*?

Peer-Assisted Learning Strategies (PALS) is one research-based approach that shows promise as an effective classwide intervention. PALS was developed by Doug and Lynn Fuchs and colleagues, in collaboration with teachers in Metro-Nashville, TN schools (see Fuchs and Fuchs 1998 for a description of this collaboration). PALS is based on the Classwide Peer Tutoring (CWPT) model developed by Delquadri and colleagues (e.g., Delquadri et al. 1986) at the Juniper Gardens Children's Project at the University of Kansas. Over the past two decades, PALS has been developed and evaluated for children in pre-K to high school, and includes procedures for both reading and mathematics.

In this chapter, Reading PALS as a classwide intervention for use in the elementary grades is described. The chapter is organized to address the five questions listed above. Specifically, the extent to which Reading PALS (a) addresses critical academic skills that align with important standards, (b) provides opportunities to differentiate instruction, (c) is supported by evidence of efficacy for diverse learners, (d) includes supports for teachers to implement PALS with fidelity, and (e) can be adapted to meet specific classroom and student needs is illustrated. The chapter concludes with a discussion of areas for future research and implications for practice.

PALS Addresses Critical Academic Skills that Align with Core Standards

During PALS, students practice critical reading skills designed to align with core curriculum and standards at each grade level (phonemic awareness, letter–sound recognition, word reading, and fluency in kindergarten and first grade; fluency and comprehension in grades 2–6). To illustrate this alignment, Table 1 provides a list of the common core standards (CCS; www.corestandards.org) that correspond with PALS activities at each grade level.

Reading PALS activities are designed to *supplement* the core curriculum, providing students with frequent opportunities to respond, engage in extended practice, and experience success in reading. PALS is implemented three times per week, for 30–45 min, depending on the grade level. Teachers typically implement PALS during their reading and language arts block or during independent reading time (although, in grades 2–6, it may also be appropriate to implement PALS as part of content-area instruction, using science or social studies texts).

Below, the PALS activities at each grade level are briefly described; more detailed descriptions can be found in other published summaries of PALS (e.g., Fuchs and Fuchs 2005; McMaster et al. 2006) and in related research articles cited below.

Kindergarten PALS (K-PALS) K-PALS comprises 72 beginning reading lessons that consist of two main parts: Teacher-Directed Sound Play and Decoding PALS. Sound Play consists of brief (3–5 min) phonemic awareness games that focus on syllables, isolating first sounds, rhyming, blending, and segmenting sounds in words, and isolating last sounds. Each Sound Play lesson aligns with a Decoding lesson. After the Sound Play activity, the teacher previews the Decoding lesson with the whole class for about 5 min, which includes introducing new sounds and sight words. Then reading partners implement the activities, which are printed on lesson sheets that partners share.

For each activity, the stronger reader in the pair is the coach first, and the weaker reader reads first, to maximize the amount of practice for the weaker reader. After each activity, the students mark a happy face printed on the lesson sheet, switch roles, and repeat the activity one time before moving to the next activity. If they have time, pairs are encouraged to repeat the activities to earn more happy faces. Decoding PALS is conducted for about 15 min. K-PALS is implemented for approximately 30 min (5 min of Sound Play, 5 min of teacher-directed lesson preview, 15 min of partner work, and transition time between activities), three times per week, and takes about 20 weeks to complete all 72 lessons.

Table 1 Reading PALS activities aligned with common core state standards at each grade level

Grade level and strand	Common core state standards (retrieved from http://www.corestandards.org/ELA-Literacy)	PALS activity
Kindergarten: phonemic awareness	CCSS.ELA-Literacy.RF.K.2 Demonstrate understanding of spoken words, syllables, and sounds (phonemes) K.2a Recognize and produce rhyming words K.2b Count, pronounce, blend, and segment syllables in spoken words K.2c Blend and segment onsets and rimes of single-syllable spoken words K.2d Isolate and pronounce the initial, medial vowel, and final sounds in three-phoneme (consonant–vowel–consonant) words	Sound play Syllables First sound identification Rhyming Guess my word (blending and segmenting) Last sound identification
Kindergarten: phonics and word recognition	CCSS.ELA-Literacy.RF.K.3 Know and apply grade-level phonics and word analysis skills in decoding words K.3a Demonstrate basic knowledge of letter–sound correspondences by producing the primary or most frequent sound for each consonant K.3b Associate the short sounds with the common spellings (graphemes) for the five major vowels K.3c Read common high-frequency words by sight (e.g., *the, of, to, you, she, my, is, are, do, does*) K.3d Distinguish between similarly spelled words by identifying the sounds of the letters that differ	Decoding PALS What sound? What word? Sound boxes Reading sentences
Kindergarten: fluency	CCSS.ELA-Literacy.RF.K.4 Read emergent-reader texts with purpose and understanding	Reading decodable books
First-grade: phonics and word recognition	CCSS.ELA-Literacy.RF.1.3 Know and apply grade-level phonics and word analysis skills in decoding words. 1.3a Know the spelling–sound correspondences for common consonant digraphs (two letters that represent one sound) 1.3b Decode regularly spelled one-syllable words 1.3c Know common vowel team conventions for representing long vowel sounds 1.3e Decode two-syllable words following basic patterns by breaking the words into syllables 1.3f Read words with inflectional endings 1.3g Recognize and read grade-appropriate irregularly spelled words	Sounds and words What sound? What word? Decoding regular words Sentences and stories
First grade: fluency	CCSS.ELA-Literacy.RF.1.4 Read with sufficient accuracy and fluency to support comprehension. 1.4b Read grade-level text orally with accuracy, appropriate rate, and expression	Partner reading
Grades 2–6: fluency	CCSS.ELA-Literacy.RF.2.4, 3.4, 4.4, 5.4 Read with sufficient accuracy and fluency to support comprehension 2.4a, 3.4a, 4.4a, 5.4a Read grade-level text with purpose and understanding 2.4a, 3.4a, 4.4a, 5.4a Read grade-level text orally with accuracy, appropriate rate, and expression 2.4a, 3.4a, 4.4a, 5.4a Use context to confirm or self-correct word recognition and understanding, rereading as necessary	Partner reading

Table 1 (continued)

Grade level and strand	Common core state standards (retrieved from http://www.corestandards.org/ELA-Literacy)	PALS activity
Grades 2–6: key ideas and details	CCSS.ELA-Literacy.RL.2.1 Ask and answer such questions as *who, what, where, when, why*, and *how* to demonstrate understanding of key details in a text CCSS.ELA-Literacy.RI.2.2 Identify the main topic of a multiparagraph text as well as the focus of specific paragraphs within the text CCSS.ELA-Literacy.RL.3.1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers CCSS.ELA-Literacy.RI.3.2 Determine the main idea of a text; recount the key details and explain how they support the main idea CCSS.ELA-Literacy.RL.4.1 Refer to details and examples in a text when explaining what the text says explicitly and when drawing inferences from the text CCSS.ELA-Literacy.RI.4.2 Determine the main idea of a text and explain how it is supported by key details; summarize the text CCSS.ELA-Literacy.RI.5.2 Determine two or more main ideas of a text and explain how they are supported by key details; summarize the text CCSS.ELA-Literacy.RL.6.1 Cite textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text	Retell and paragraph shrinking

PALS Peer-Assisted Learning Strategies

Students learn specific prompts and correction procedures for each of the Decoding activities. First, “What Sound?” features consonants and vowels printed at the top of the lesson sheet. The coach points to each letter and says, “What sound?” and the reader says each sound. If the reader makes a mistake (e.g., on the/m/sound), the coach says, “Stop, that sound is/mmm/. What sound? Good, read that line again.” Second, “What Word?” features common sight words (such as “the,” “was,” and “is”) printed below the “What Sound?” letters on the lesson sheet. The coach points to each word and says, “What word?” and the reader reads each word, receiving corrective feedback as needed from the coach. The third activity features decodable words, consisting of the sounds practiced in “What Sound,” printed in “Sound Boxes” below the sight words. The coach prompts the reader to “Read the word slowly;” the reader points to and says each sound without stopping between sounds. Then the coach says, “Sing it and read it.” The reader blends the sounds together and then reads the word quickly, with the coach providing corrective feedback. Finally, the students read a sentence made up of decodable and sight words practiced in the lesson.

About 10 weeks into K-PALS, when students can read a sufficient number of decodable and sight words, the teacher selects brief decodable books for students to read in a partner reading format. The stronger reader in the pair reads the book first, pointing to each word. Then, the weaker reader reads the same book. Each partner reads the book twice (for a total of four readings), and then trades it in for a new book. This activity takes about 5 min.

First-Grade PALS First-Grade PALS includes approximately 70 lessons that consist of two main parts: Sounds and Words and Partner Reading. For Sounds and Words, the teacher introduces new sounds and sight words. Then, PALS pairs implement the activities, which are printed on lesson sheets that partners share. For each activity, the stronger reader is the coach first, and the weaker reader reads first, to maximize the amount of practice for the weaker reader. After each activity, the students mark a happy face printed on the lesson sheet, mark five points on a point sheet, switch roles, and repeat the activity one time before moving to the next activity. If they have time, pairs are encouraged to repeat the activities to earn additional points. Following Sounds and Words, which takes about 30 min, pairs conduct Partner Reading using children's literature for 10 min. Altogether, First-Grade PALS is implemented for approximately 40 min per session, three times per week, and takes about 20 weeks to complete.

As in K-PALS, students learn specific prompts and correction procedures for each of the Sounds and Words activities. The "What Sound?" activity is similar to that in K-PALS, although more sounds are practiced in a given session. After "What Sound?" students read decodable words that are presented using the "ball and arrow" method: the reader puts his or her finger on the ball printed at the beginning of the word. When the coach says, "Sound it out," the reader points to and says each sound, sliding his or her finger along the arrow under the word while blending the sounds together. When the coach says, "Read it fast," the reader reads the word quickly, receiving corrective feedback from the coach. After this activity, the students read common sight words. The sight words are chunked into phrases, such as "he saw the" to encourage fluent reading. Sight words are followed by brief stories made up of practiced decodable and sight words.

Following the Sounds and Words activities described above, the students engage in a repeated reading activity designed to build automatic word recognition. Early in the program, a "Speed Game" is conducted with the sight words, in which each reader has the opportunity to read the sight words three times, and attempt to read more words each time. Later in the program, the "Speed Game" is conducted with the PALS stories. When students increase the number of words read on the second or third try, they mark stars on their own personal "Star Charts," and earn prizes (e.g., PALS bookmarks) when they complete their charts.

After the Speed Game, pairs conduct Partner Reading using books selected at the weaker reader's level. This time the stronger reader in the pair reads first, one sentence at a time and pointing to each word, to provide a fluent model for the weaker reader. Then, the weaker reader reads the same sentence, pointing to each word. They proceed through the entire book in this manner, and then read it again, this time with the weaker reader reading first. The pair reads the book at least four times, and then trades it in for a new book.

PALS for Grades 2–6 PALS for Grades 2–6 focuses on fluency and comprehension skills and includes four main components: Partner Reading, Retell, Paragraph Shrinking, and Prediction Relay. The teacher introduces, models, and provides practice in each activity across 12 lessons. Then PALS pairs implement the activities for approximately 35 min per session, three times per week. Researchers recommend that PALS be implemented for a minimum of 15–18 weeks, which is the typical duration of PALS implementation in research (Fuchs et al. 1997; Fuchs et al. 2010).

Teachers select texts that are at the weaker reader's instructional level (i.e., students should read at approximately 90 % of the words correctly). As described in the next section, the gap between pairs is intended to not be too wide; thus, the text should also be appropriate for the stronger reader (at his/her independent reading level, but not too easy). Partners share one copy of the text. During each PALS session, Partner Reading is first implemented for 10 min. The teacher identifies the stronger reader as the "first reader" and the weaker reader as the "second reader." The stronger reader reads for 5 min to provide a fluent model. Then, the weaker reader reads the same text for 5 min. The coach (whoever is not reading) follows along and provides corrective feedback. If the reader makes a word-reading error, the coach points to the word and says, "Check it." The reader can either self-correct or ask for help, in which case the coach says, "That word is ___. What word?" The reader repeats the word and then rereads the sentence. The coach marks one point for each correctly read sentence. After the weaker reader reads, the stronger reader prompts a Retell of the story, by asking, "What happened first, what happened next," and so on, for 2–3 min. The pair marks up to ten points for this Retell.

Next, Paragraph Shrinking is implemented for 10 min. The stronger reader reads first. After each paragraph, the coach prompts the reader to, "Name the most important who or what," "Tell the most important thing about the who or what," and "Say the main idea in ten words or less." The reader earns one point for each response. If the reader is stuck, the coach helps by asking the reader to skim the paragraph, providing a hint, and providing an answer if needed. After 5 min, the weaker reader reads *new* text, for 5 min, following the same procedures.

Last, Prediction Relay is implemented for 10 min. The stronger reader again reads first. This time, the coach asks the reader, "What do you predict will happen next?" and the reader makes a prediction. The coach says, "Read half a page," and the reader reads. Then, the coach asks, "Did your prediction come true?" and the reader confirms or disconfirms the prediction. The reader earns one point for each response. This process continues for 5 min, and then the weaker reader reads *new* text, for 5 min, following the same procedures.

PALS Provides Opportunities to Differentiate Instruction

PALS was designed to accommodate academic diversity in general education classrooms (Fuchs et al. 1997), and can include *all* students in the classroom. Because students work in pairs during PALS, the teacher can differentiate instructional materials, pacing, and feedback to target individual students' learning needs. Below, PALS components that allow for this differentiation and are common to all grade levels are described.

PALS partners In PALS, higher-performing readers are paired with lower-performing readers to practice the reading skills described above. Typically, the teacher rank-orders all of the students in the classroom based on reading skill level, using recent progress-monitoring data. Then, the teacher splits the rank-ordered list in half, and pairs the top student from the top half with the top student from the bottom half, and so on, until all students are paired. Thus, each pair includes a stronger and a weaker reader, but the discrepancy is not so large as to introduce frustration between partners or difficulty in selecting appropriate reading materials. The teacher also takes social and behavioral skills into account, and adjusts pairs accordingly. Pairs work together for about a month, and then the teacher re-ranks the class (again using recent data), and creates new pairs, allowing students the opportunity to work with a variety of peers.

Occasionally, teachers modify the pairing procedure for students for whom the above process seems inappropriate. For example, students with very low decoding skills who struggle to read connected text may work one-on-one or in small groups with an adult to practice needed reading skills. Alternatively, they may serve as cross-age tutors in kindergarten or first-grade PALS classrooms, allowing them to practice needed basic reading skills while helping younger students. Students with very advanced reading skills may be paired together to read more challenging texts. Students with behavior problems that interfere with productive partner relationships may be paired with an adult or with an "invisible partner." While such approaches may be effective solutions, it is important to keep in mind that the standard partnering approach described above is ideal for PALS implementation. Teachers are thus encouraged to find ways to work toward having all students work in productive PALS pairs.

Structured reciprocal peer tutoring During PALS, partners take turns being coach and reader. The teacher provides explicit training in these roles and PALS procedures through brief scripted lessons. During this training, students learn structured prompts and corrective feedback that the coaches use while the readers read. In each PALS session, students take turns as both coach and reader, allowing the weaker reader to observe a more fluent reading model and practice critical skills with immediate feedback, and also providing the stronger reader with the opportunity to practice and refine critical reading skills. The highly structured, consistent procedures allow students to conduct PALS independently, and foster high engagement (Fuchs et al. 1997).

Differentiated reading material Because students work in pairs during PALS, the teacher can select material that is appropriate for the weaker reader in each pair. Given the relatively small discrepancy between partners (as described above), the material should also be appropriate for the stronger reader. In kindergarten and first grade, the whole class typically proceeds through the lessons at the same pace; however, the teacher can decide to place specific pairs on earlier (or later) lessons, according to individual needs. In grades 2–6, PALS is conducted with classroom reading material (e.g., children's literature, leveled readers, basal readers, or content-area texts). The teacher selects texts at the instructional level of the weaker reader, as determined by the general rule that a student should read approximately 90 % of the words correctly, or by leveling systems used in the school curriculum. Often, the teacher allows student choice, by providing an array of texts of different genres at a given level for pairs to select. Students read multiple texts across the school year during PALS.

Teacher monitoring and positive reinforcement Because students work in pairs during PALS, the teacher is able to walk around the classroom to ensure students are following procedures and reading materials are appropriate, and to provide individualized feedback as needed. Teachers can award points to pairs for engaging in specific reading behaviors (e.g., reading with expression) and social behaviors (e.g., providing help to the reader). PALS pairs also award themselves points for completing each activity. Points are recorded on a point sheet that is shared between two partners. Each pair is placed on one of two teams. At the end of the week, pairs report their weekly point total, and points are added up for each team. Both teams are recognized for their hard work, and the winning team celebrates (e.g., by taking a bow).

PALS has Evidence of Efficacy for Diverse Learners in General Education Classrooms

Below, research demonstrating PALS efficacy as a classwide intervention at each elementary grade level is briefly reviewed, highlighting evidence of efficacy for diverse learners with and without disabilities in kindergarten, first grade, and grades 2–6. Research designed specifically to examine students' responsiveness to PALS is also described.

Kindergarten PALS In the late 1990s, PALS researchers worked with classroom teachers to design and test K-PALS across several years of research (see Fuchs and Fuchs 2005), which included contrasts of versions of PALS with various combinations of teacher-directed and peer-mediated phonological awareness (PA) and decoding activities. In the culminating randomized field trial (Fuchs et al. 2001a, b), 33 kindergarten teachers in eight urban schools were assigned randomly within school to three groups: PA only, PA + Decoding PALS, and business-as-usual control. Teachers in the two treatment groups implemented intervention for 20 weeks. They were asked to adhere to their assigned condition, and were observed twice to confirm fidelity of implementation, using detailed checklists of teacher and student behaviors prescribed by the K-PALS training and manual.

Pre- and posttest beginning reading data were collected from low-, average-, and high-performing children in each classroom. Students of each learner type who received PA + Decoding PALS outperformed both those who received PA only and controls on PA (segmenting and blending) and alphabetic (letter–sound identification, word attack, word identification, and spelling) measures. Effect sizes (ESs) favoring PA + Decoding PAS versus controls on PA measures ranged from 0.45 to 1.27 (low-performing readers), 1.97–2.10 (average-performing readers), and 1.07–1.21 (high-performing readers). ESs favoring PA + Decoding PAS versus controls on alphabetic measures ranged from 0.28 to 1.28 (low-performing readers), 0.73 to 1.42 (average-performing readers), and 0.08 to 0.90 (high-performing readers).

Following the large-scale efficacy trial, researchers have examined K-PALS effects for subgroups of diverse learners. Fuchs et al. (2002) disaggregated data for kindergartners with disabilities from the study described above. Twenty-four participants in the larger K-PALS students were identified as having disabilities based on current Individualized Education Programs (IEPs), with the majority (84 %) identified as having speech or language impairments. There were no statistically significant pretreatment differences between K-PALS and control groups in terms of demographic or reading-related variables. On average, K-PALS students with disabilities outperformed comparison students with disabilities on measures of PA, letter–sound recognition, and word attack (ESs = 0.39–0.69).

Rafdal et al. (2011) replicated these findings by disaggregating data for kindergartners with disabilities from a multisite K-PALS study (described below; Fuchs et al. 2010). Rafdal et al. (2011) included 89 students with IEPs who scored within three standard deviations of the mean on pretest reading measures (excluding students with severe cognitive disabilities or highly advanced reading skills). Again, the majority of participants were identified as having speech/language disorders, with the remainder identified as having learning disabilities, emotional/behavioral disorders, developmental cognitive delay, or attention deficits. There were no statistically significant pretreatment differences between K-PALS and control groups on demographic variables. Multivariate analyses of covariance (using pretest reading performance as the covariate) revealed that K-PALS students with disabilities outperformed controls on word attack, spelling, and oral reading (ESs = 0.31–0.51).

Whereas both studies above suggested that K-PALS students with disabilities outperformed controls on important beginning reading measures, it is critical to note that, in both studies, some students with disabilities made very little (or no) progress. In other words, K-PALS was effective for some, but not all, kindergarten students with disabilities. This finding underscores the idea that, for students with disabilities, it is important to determine whether practices deemed generally effective are appropriate for meeting specific individual needs.

Researchers have also disaggregated K-PALS data for English learners (ELs). McMaster et al. (2008) examined effects of K-PALS for ELs from a variety of cultural and linguistic backgrounds in urban Midwestern classrooms. They found that ELs who received K-PALS outperformed ELs in control classrooms (using data from the Fuchs et al. 2010 investigation mentioned above) with ESs ranging from 0.65 to 0.69 on PA measures, 0.58 on letter–sound identification, and 0.22 on word attack. Further, K-PALS appeared to reduce proportions of ELs who were nonresponsive to core instruction (e.g., on letter–sound identification, 5 % of ELs were identified as nonresponsive compared to 35 % on control ELs; a chi-square analysis indicated that the difference in proportions was statistically significant). These results support the use of K-PALS as a supplement to core instruction for ELs in kindergarten classrooms.

First-Grade PALS First-Grade PALS efficacy research has proceeded in a similar fashion to K-PALS research, but with a somewhat different focus. Specifically, researchers contrasted a decoding-only version of First-Grade PALS with a fluency-building version that incorporated repeated reading of sight words and connected text in a "Speed Game" format (PALS + Fluency; Fuchs et al. 2001c). In a randomized field trial, 33 first-grade teachers in title I and non-title I schools were assigned randomly within school to one of the two PALS versions or to control. Treatment teachers implemented PALS for 22 weeks. Low-, average-, and high-performing students in both PALS groups reliably outperformed their counterparts in control classrooms on measures of PA (ES = 0.50), word-level reading skills (ES = 0.21), and fluency (ES = 0.19; Fuchs et al. 2013). There were no reliable differences between the two PALS versions, suggesting that either version can be used to boost word reading and fluency skills.

Researchers have also demonstrated a benefit of First-Grade PALS for ELs. Calhoon et al. (2007) examined the effects of First-Grade PALS on the reading achievement of students in two-way bilingual immersion classrooms (classrooms that allocated equal time to both English and Spanish) in the southwest, on the border of Mexico. They assigned six classrooms from three schools randomly to PALS or to a control condition. A majority of student participants were Hispanic, identified either as English proficient students or ELs. Findings indicated statistically significant effects of PALS on growth in PA (ES =0.53), decoding (ES =0.50), and oral reading skills (ES =0.51) for both English-proficient students and ELs, again supporting its use in diverse classrooms

PALS for Grades 2–6 Reading PALS for Grades 2–6 was initially developed in the mid-1990s. In a randomized control trial (Fuchs et al. 1997), 12 urban and suburban schools were stratified on socioeconomic status and reading achievement levels, and assigned randomly to experimental and control groups. Forty teachers in grades 2 through 6 participated; 20 implemented PALS for 15 weeks, and 20 served as controls. Pre- and posttest oral reading and reading comprehension data were collected from three students from each classroom representing three learner types (one low-achieving and one average-achieving based on teacher judgment, and one with LD based on district disability certification).

Students from PALS classrooms outperformed those in controls on number of words read correctly in 3 min (d =0.22), number of comprehension questions answered correctly (d =0.55), and number of correct maze choices on the Comprehensive Reading Assessment Battery (CRAB; d =0.56). PALS effects were not moderated by learner type, suggesting that PALS could be used successfully in classrooms in which students with LD were included. An important caveat, however, was that 20 % of PALS students with LD did *not* make marked improvement in reading, indicating that some students may be in need of more intensive intervention (Fuchs et al. 1997).

PALS has also been demonstrated to be effective for native Spanish-speaking students with and without LD in grades 3 through 6. In a randomized control trial conducted in South Texas, Saenz et al. (2005) stratified 12 classrooms based on grade level and school, and assigned them randomly to PALS or control. Student participants included 132 ELs with LD and their low-, average-, and high-achieving classmates. PALS students outperformed controls on a measure of reading comprehension regardless of learner type, with ESs of 0.60 for words read correctly in 3 min, 1.02 for number of comprehension questions answered correctly, and 0.40 for correct maze choices (all on the CRAB; Fuchs et al. 1989). These findings provide further evidence of the efficacy of PALS in improving reading outcomes for diverse learners.

Responsiveness to PALS Despite the general benefits of PALS, approximately 20 % of low-achieving students without disabilities (Al Otaiba and Fuchs 2006) and more than 50 % of students with disabilities (Fuchs et al. 2002) have not responded to PALS, as measured by growth on tests of PA, decoding, and word recognition. Al Otaiba and Fuchs (2006) attempted to describe characteristics of PALS “nonresponders.” They conducted a study with 104 children who had participated in PALS in kindergarten only, first grade only, both kindergarten and first grade, or neither grade. These children were classified as “always responsive,” “sometimes responsive,” or “nonresponsive” based on kindergarten and first-grade reading achievement.

Using multivariate analysis of variance and discriminant function analysis, Al Otaiba and Fuchs (2006) found that a combination of rapid naming speed, vocabulary, working memory, behavior, and amount of intervention (PALS) received over the 2 years predicted 82 % of non-responsive, 30 % of sometimes responsive, and 84 % of always responsive students. A subset of these students (n =50) were tested again at the end of their third-grade year. Of those who had participated in PALS, all but one were receiving special education services in reading. These findings suggest the need to provide more intensive intervention to young children who are unresponsive to classwide intervention as early as possible.

McMaster et al. (2005) attempted to address student nonresponsiveness by identifying students at risk of early reading failure and monitoring their response to First-Grade PALS for the first 7 weeks of implementation. Students who did not make sufficient progress (based on a dual discrepancy of 0.50 standard deviation (SD) below average-performing readers' levels and slopes on word-reading progress measures) were assigned randomly to continue receiving PALS, to receive a modified version of PALS that was implemented in a classwide setting but tailored to individual needs, or to receive one-to-one adult tutoring outside of the classroom, instead of PALS.

Analyses of covariance indicated that neither modified PALS nor tutoring produced statistically significantly greater performance on measures of PA, alphabetic knowledge, word reading, or oral reading fluency than did the regular PALS intervention. Further analysis revealed that 81 % of students who remained in PALS, 80 % of students who received modified PALS, and 50 % of students who received tutoring continued to qualify as "nonresponders" based on the dual discrepancy criteria above. Chi-square analysis showed that the proportion of nonresponders to tutoring was statistically significantly lower than the proportion of nonresponders to PALS. In other words, though there were no differential effects among the three groups *on average,* examination of individual students' responsiveness revealed that more students experienced benefits from tutoring than from the other PALS interventions. These findings underscore the importance of monitoring student responsiveness to classwide intervention, and making timely changes when data indicate the need. It also remains clear that some students will likely require more intensive, individualized intervention than can be offered in a classwide setting.

Professional Development and Support Improves Fidelity of PALS Implementation

A major assumption of RTI is that each level of instruction and intervention is implemented with *fidelity*—that is, that instructional programs are implemented as intended by the developers (Noell and Gansle 2006). Without evidence of implementation fidelity, accurate decisions about individual students' responsiveness to high-quality instruction and need for more intensive intervention may be seriously compromised. Recent work focused on effects of PALS when implemented at scale have included examinations of the types of professional development and support needed to promote teachers' fidelity of PALS implementation, the extent to which PALS fidelity mediates student achievement, and ways to provide cost-effective support within districts that attempt to adopt and sustain PALS as a supplement to core instruction.

Scaling Up PALS In 2004, the Institute of Education Sciences (IES; US Department of Education) awarded a 5-year Scale-Up Evaluation Grant to D. Fuchs and colleagues at Vanderbilt University, The University of Texas Pan American, and the University of Minnesota (see Fuchs et al. 2010). The purpose of the project was to "determine whether or not an intervention is effective when it is implemented…across a variety of conditions" (IES 2010, p. 9). In this project, researchers examined PALS effects on student reading achievement across different student populations and types of schools, including in Nashville, TN, where schools had considerable familiarity with and history of using PALS; Minnesota, where schools had some history of using PALS; and South Texas, where schools had little or no history of using PALS. The project included efforts to scale up K-PALS and PALS in grades 2–5.

The K-PALS study (Fuchs et al. 2010; Stein et al. 2008) focused on the level of support that teachers would need to implement K-PALS effectively and with fidelity (i.e., the accuracy with which they implemented K-PALS, as measured by a detailed observational checklist of all K-PALS components). Across the three sites, teachers were assigned randomly to (1) Control, (2) Workshop Only (teachers attended a one-day workshop and then implemented PALS on their own), (3) Workshop + Boosters (teachers attended the workshop plus two to three 1-h problem-solving sessions with PALS researchers and other teachers), or (4) Workshop + Boosters + Helper (teachers received weekly onsite assistance from a research assistant).

After 18 weeks, teachers in all sites who received Workshop + Boosters or Workshop + Boosters + Mentor implemented K-PALS with greater fidelity (88–93 % components implemented correctly) than did teachers in the Workshop Only group (80 % components implemented correctly; McMaster et al. 2009). K-PALS students outperformed controls on measures of PA, regardless of site or level of support (ESs = 0.29–0.42). Effects varied by site, however, on measures of letter–sound identification, word reading, and fluency (McMaster et al. 2009). In Tennessee, effects were generally strong, though not as strong as in previous research (ESs = 0.27 to 0.93). In Minnesota, effects were weak to moderate (ESs = 0.07 to 0.49). In Texas, effects were negligible or favored controls (ESs = -0.09 to 0.32), except on letter–sound recognition (ESs = 0.12 to 0.33). Effects also varied based on level of teacher support, with Workshop + Boosters adding value above Workshop Only and Workshop + Booster + Helper groups on letter–sound (e.g., ESs = 0.11 to 0.30) and word attack (ESs = 0.22 to 0.28) outcomes.

Varied outcomes across sites may be attributed to the fact that Tennessee had the most K-PALS experience and resources. Varied outcomes across levels of support may be attributed to differences in fidelity. As described above, teachers in the Booster and Helper groups implemented PALS with greater fidelity than did teachers in the Workshop Only group (McMaster et al. 2009; see also Stein et al. 2008). Yet, it appeared that the Boosters added value to K-PALS effects, whereas the Helper condition did not. A possible explanation for this finding is that Boosters provided teachers with the support needed to implement K-PALS with sufficient fidelity to affect student outcomes. However, in the Workshop + Booster + Helper group, teachers reported feeling that they had to adhere to K-PALS procedures rather rigidly, because they knew that the Helper would provide feedback focused on fidelity. Though the Helper was not intended to play an evaluative role, teachers perceived it this way. Thus, researchers concluded that *too* much emphasis on fidelity (such as that conveyed in the Helper condition) could lead to "a kind of 'K-PALS fatigue,' where teachers [feel] confined to continue the program as designed" (Kearns et al. 2010, p. 336). This finding does not necessarily suggest that onsite technical assistance cannot be effective, but raises questions regarding how to best balance an emphasis on fidelity with some degree of flexibility of implementation.

Sustaining PALS In addition to supporting teachers' initial adoption of a research-based classwide intervention, it is critical to determine how to support teachers' sustained use of the intervention over time. Kearns et al. (2010) explored factors that predicted teachers' sustained use of K-PALS 1 year after they first adopted it for use in their classrooms. Teachers who had been involved in the above scaling-up research were followed up with interviews that asked whether they were continuing to use K-PALS and what influenced this decision. The strongest predictors of teachers' reported sustained use were their perceptions of K-PALS effectiveness for their students, and the type and degree of external support provided. Consistent with the scaling-up results, teachers who had received "Helper" support were *less likely* to report sustaining K-PALS than were those who received "Booster" support. This finding could reflect the "PALS fatigue" suggested above, or perhaps too much dependency on the Helper, such that teachers were reluctant to implement PALS independently (Kearns et al. 2010). Findings suggest the need for ongoing support to sustain PALS that includes some way of promoting teachers' sense of ownership and confidence in their ability to implement PALS independently.

Given that support is likely needed to promote teachers' successful sustained use of PALS, an important question is *how* this support might be delivered in a cost-effective way in schools. Thus, McMaster et al. (2013) compared effects of university-provided support versus district support on teachers' fidelity of K-PALS implementation and students' reading achievement. In a district that decided to adopt K-PALS as part of their literacy program following participation in the scaling-up research, 16 teachers who were new to K-PALS were assigned randomly to receive ongoing support from a university expert or from experienced K-PALS teachers within the district. Support consisted of Booster sessions similar to those implemented in the scaling-up study.

In both university- and district-supported classrooms, K-PALS teachers' students reliably outperformed historical controls on beginning reading measures (d =0.24–1.29). There were no statistically significant differences between university- and district-supported groups with respect to teachers' implementation fidelity or students' reading achievement. Thus, it appears that providing support within districts (in which there are experienced K-PALS teachers) may be an effective way of providing support to teachers, and may be more feasible than providing university support, given limited resources in schools. Important caveats include that the district in this study had an ongoing relationship with K-PALS researchers, that some teachers in the district had been part of a year-long K-PALS study in which they received university support, and the district continued to receive limited support (primarily in terms of training). Also, the study was conducted in a suburban district with strong administrative support for K-PALS, teacher buy-in to using K-PALS, and relatively stable student enrollments. Findings may have been quite different had the study been conducted under different district conditions.

PALS May Be Adapted to Meet Specific Classroom and Student Needs

Whereas findings of the scaling-up K-PALS work highlight the importance of promoting and supporting teachers' fidelity of implementation, the relation between fidelity and student outcomes is not entirely straightforward. As mentioned, it appeared that gains in fidelity achieved through the most intensive (Helper) level of teacher support did *not* translate to improved student achievement outcomes (Fuchs et al. 2010). This finding is somewhat counterintuitive given the current general emphasis on the importance of fidelity in implementing scientific, research-based instruction (e.g., Gersten et al. 2009).

Another unexpected outcome also occurred: The effects observed in the scaling-up K-PALS study were relatively modest compared to findings from earlier research. Closer investigation revealed that control students in the scaling-up study achieved at notably higher levels than controls in earlier K-PALS research (Lemons et al. 2014), suggesting that kindergarten reading instruction was generally stronger than it had been a decade ago. In other words, it appears K-PALS was pitted against a stronger control, perhaps due to changes in kindergarten reading instruction spurred from the release of the National Reading Panel report (National Institute of Child Health and Human Development 2000) and the advent of Reading First (2002).

The collective findings from the scaling-up K-PALS work led researchers to conclude that the impact of a research-based intervention may vary across both time and educational setting, and that it may be important to build in flexibility to adapt PALS to better fit varied contexts (Fuchs et al. 2010). Thus, when the focus of the scaling up project turned to upper elementary grades, researchers determined that, while it may be important to maintain teachers' fidelity to core PALS elements, encouraging some flexibility in their use of PALS might increase its robustness across a variety of conditions.

Teachers in grades 2–5 participated in the study for 2 years. In year 1, teachers were assigned randomly to PALS or Control. All PALS teachers were asked to implement "Top-Down PALS"—in other words, to use it exactly as described in the manual—to ensure that they became proficient in implementing all PALS components. Then, in year 2, PALS teachers chose to implement either "Top-Down" or "Bottom-Up" PALS. Control teachers continued to serve as controls. Bottom-Up PALS teachers were asked to implement core elements of PALS that have strong research support—10 min of Partner Reading and 10 min of Paragraph Shrinking, along with a motivational component–for at least 35 min per session for 48 sessions. They were also strongly encouraged to customize noncore elements, which could include making minor tweaks, such as changing the point system; big changes, such as replacing Retell and Prediction Relay with new activities; and flexible changes, by varying activities over time. Bottom-Up teachers developed their own customizations, but were given support from researchers to develop materials and refine their activities as needed. All Bottom-Up teachers opted to make big changes, incorporating a wide variety of vocabulary, comprehension, and writing-related strategies.

Results of the study revealed that, together, Top-Down and Bottom-Up PALS students made reliably greater reading gains than controls (ES =0.25). However, when separated into the two groups, the effect of Top-Down PALS compared to controls was small and not statistically significant (ES =0.15), whereas the effect of Bottom-Up PALS compared to controls was moderate and statistically significant (ES =0.34). Further, Bottom-Up PALS students made small, but reliably greater reading gains than Top-Down PALS students (ES =0.19). Top-Down and Bottom-Up teachers did not differ on other important variables, such as level of education or experience, general teaching effectiveness, or overall PALS fidelity, suggesting that differences in student performance can be attributed to Bottom-Up PALS (Fuchs et al. 2010).

These findings support the notion that some degree of flexibility to adapt classwide intervention to meet classroom- or student-specific needs may have a positive effect on student outcomes. Note, however, that these results were obtained under three important conditions: (a) teachers had implemented Top-Down PALS with fidelity for 1 year, thus gaining proficiency in all PALS components before making adaptations, (b) Bottom-Up PALS teachers implemented *core* PALS components with fidelity, and only adapted noncore components, and (c) Bottom-Up PALS teachers received ongoing support in the form of Boosters from PALS experts. There is currently no empirical evidence that adaptations implemented outside these parameters would have similarly positive effects on student achievement.

Future PALS Research

PALS has extensive support as a scientific, research-based classwide intervention. Additional research would continue to improve our understanding of the conditions under which PALS is most effective, as well as for whom PALS is most beneficial. In the context of RTI, an important question is whether PALS enhances the quality of core instruction in RTI models, such that proportions of students in need of more intensive levels of intervention are reduced. Such a question is especially important in schools with high proportions of students who are at risk, given limited resources to deliver more intensive intervention to large numbers of children. In addition, more research is needed to understand the long-term benefits of PALS, as well as the best ways to help students for whom PALS is not sufficiently beneficial.

Other important questions relate to the sustainability of PALS. Whereas researchers have begun to investigate factors related to teachers' sustained use of PALS, more work is needed to understand whether and how teachers continue to use PALS over multiple years, and how to best support their sustained use. Further work is needed to understand how teachers adapt PALS over time, and conditions under which their adaptations are more or less effective. For schools that adopt PALS at multiple grade levels, it is also important to investigate the cumulative impact on academic achievement of students who participate in PALS across multiple years and in multiple content areas. It would also be very useful to develop guidelines for determining when PALS is not sufficient to promote individual students' reading growth, indicating the need for more and/or different reading instruction.

Implications for Practice

Given the extent of evidence supporting the efficacy of PALS for diverse learners, it is recommended that schools in need of improving core instruction consider adopting PALS as a classwide intervention. In doing so, it may be useful to consider the five questions introduced at the beginning, and addressed throughout, this chapter.

Table 2 Using PALS in practice: considerations, evidence, and recommendations for practitioners

Considerations for adopting: classwide intervention	PALS support	Recommendations for practitioners
Does the intervention address critical academic skills that align with curriculum and standards?	Reading PALS activities align with grade-level standards such as the CCSS	Determine whether the skills addressed in PALS align with specific standards used by the school and district
Does the intervention provide opportunities to differentiate instruction?	Reading PALS allows for differentiation through: PALS partners Structured reciprocal peer tutoring Differentiated reading material Teacher monitoring, and positive reinforcement	For reading PALS to be successful, it is critical to implement PALS with fidelity and to: Pair stronger readers with weaker readers Teach effective coaching behaviors Select reading materials at the weaker reader's level Monitor students' reading *and* social behaviors Provide corrective feedback and positive reinforcement
Does the intervention have evidence of efficacy for students similar to those in the classroom in which it is to be used?	Evidence supports PALS effectiveness for a wide range of learners, including students with disabilities (Fuchs et al. 2002; Rafdal et al. 2011), and English learners (Calhoon et al. 2007; McMaster et al. 2008; Saenz et al. 2005)	Examine participants sections of relevant PALS studies to determine whether participants were similar to students in your classroom
Criteria: classwide intervention should…	*PALS support*	*Recommendations for practitioners*
What types of professional development and supports are needed to ensure fidelity of implementation?	PALS is a manualized intervention with clearly laid-out scripts and materials for implementation PALS fidelity may improve with ongoing support (Fuchs et al. 2010; Stein et al. 2008) Support may be provided by university or district experts (McMaster et al. in press)	Obtain PALS manual and training (see kc.vanderbilt.edu/pals/) Assemble PALS materials (including appropriate texts) Schedule adequate time to implement PALS Implement PALS with fidelity Access ongoing support from PALS trainers and/or experienced PALS teachers locally if available
Can the intervention be adapted to meet specific classroom and student needs?	Evidence suggests that when teachers (a) implement core PALS components with fidelity and (b) adapt noncore PALS components to suit their students' needs, student achievement improves (Fuchs et al. 2010)	Implement PALS as prescribed until both teacher and students are proficient in all activities Implement core PALS components with fidelity As needed, adapt noncore PALS components to better fit classroom or individual students' needs Implement adaptations systematically and with fidelity Monitor student progress using appropriate assessments to determine effectiveness

PALS Peer-Assisted Learning Strategies

Table 2 provides a summary of these questions, evidence of the extent to which PALS addresses these questions, and recommendations for practice. More information about PALS research, as well as professional development and materials, can be found at http://kc.vanderbilt.edu/pals/.

Conclusion

In this chapter, PALS is described as one approach to delivering scientific, research-based classwide intervention in reading that can be used in general education elementary classrooms and within multi-tiered systems of support (PALS is integrated into core instruction). PALS components at each grade level are described, and research supporting the efficacy of PALS to improve reading outcomes for diverse learners is reviewed. As with any research-based classwide intervention, PALS has the promise to benefit many, but not all children. Thus, the authors emphasize the importance of implementing core components with fidelity to ensure that students have the opportunity to respond to high-quality PALS implementation. At the same time, it is critical to monitor the effects of PALS and provide different interventions for students do not respond adequately to PALS alone.

References

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? *Journal of Learning Disabilities, 39,* 414–431.

Al Otaiba, S., Kosanovich-Grek, M. L., Torgesen, J. K., Hassler, L., & Wahl, M. (2005). Reviewing core kindergarten and first-grade reading programs in light of No Child Left Behind: An exploratory study. *Reading & Writing Quarterly, 21,* 377–400.

Burns, M. K. (2010). Response to intervention research: Is the sum of the parts as great as the whole. *Perspectives on language and literacy, 36*(2), 13–15.

Calhoon, M. B., Al Otaiba, S., Cihak, D., King, A., & Avalos, A. (2007). Effects of a peer-mediated program on reading skill acquisition for two-way bilingual first-grade classrooms. *Learning Disability Quarterly, 30,* 169–184.

Delquadri, J., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional Children, 52,* 535–542.

Elementary and Secondary Education Act 20 U.S.C. § 6301. (2002)

Fuchs, D., & Fuchs, L. S. (1998). Researchers and teachers working together to adapt instruction for diverse learners. *Learning Disabilities Research and Practice, 13,* 126–37.

Fuchs, D., & Fuchs, L. S. (2005). Peer-assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal of Special Education, 39,* 34–44.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research, 83,* 103–111.

Fuchs, D., Fuchs, L. S., Mathes, P. H., & Simmons, D. C. (1997). Peer-assisted strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34,* 174–206.

Fuchs, D., Fuchs, L. S., Al Otaiba, S., Thompson, A., Yen, L., McMaster, K. N., Svenson, E., & Yang, N. J. (2001a). K-PALS: Helping kindergartners with reading readiness: Teachers and researchers in partnerships. *TEACHING Exceptional Children, 33,* 76–80.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N. J., Braun, M., & O'Connor, R. E. (2001b). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93,* 251–267.

Fuchs, D., Fuchs, L., Yen, L., McMaster, K., Svenson, E., Yang, N., Young, C., Morgan, P., Gilbert, T., Jaspers, J., Jernigan, M., Yoon, E., & King, S. (2001c). Developing first-grade reading fluency through peer mediation. *TEACHING Exceptional Children, 34,* 90–93.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N. J., Braun, M., & O'Connor, R. E. (2002). Exploring the importance of reading programs for kindergartners with disabilities in mainstream classrooms. *Exceptional Children, 68,* 295–311.

Fuchs, D., McMaster, K., Saenz, L., Kearns, D., Fuchs, L., Yen, L., Compton, C., Lemons, C., Zhang, W., & Schatschneider, C. (2010). Bringing educational innovation to scale: Top-down, bottom-up, or a third way? Presented at the IES Conference, Washington, DC.

Fuchs, D., Fuchs, L. S., Toste, J. R., McMaster, K. L., Svenson, E., & Thompson, A. (2013). First Grade Peer-Assisted Learning Strategies (PALS): Does fluency-building enhance reading performance? Manuscript in preparation.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention (RtI) and multi-tier intervention in the primary grades*. Washington, DC: U.S. Department of Education. (IES Practice Guide).

Individuals With Disabilities Education Act, 20 U.S.C. § 1400. (2004).

Institute for Education Sciences. (2010). Request for grant applications: Education research grants. http://ies.ed.gov/funding/.

Kearns, D. M., Fuchs, D., McMaster, K. L., Saenz, L., Fuchs, L. S., Yen, L., Meyers, C., Stein, M., Compton, D., Berends, M., & Smith, T. M. (2010). Factors contributing to teachers' sustained use of kindergarten peer-assisted learning strategies. *Journal of Research on Educational Effectiveness, 3,* 315–342.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-Based Practices in a Changing World Reconsidering the Counterfactual in Education Research. Educational Researcher, 0013189X14539189.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71,* 445.

McMaster, K. L., Fuchs, D., & Fuchs, L. S. (2006). Research on peer-assisted learning strategies: The promise and limitations of peer-mediated instruction. *Reading & Writing Quarterly, 22,* 5–25.

McMaster, K. L., Kung, S., Han, I., & Cao, M. (2008). Peer-assisted learning strategies: A "tier 1" approach to promoting English learners' response to intervention. *Exceptional Children, 74,* 194–214.

McMaster, K., Saenz, L., Fuchs, D., Berends, M., Yen, L, & Kearns, D. (2009). *Scaling up peer-assisted learning strategies: Findings from three years of research*. Paper presented at the 2009 council for exceptional children annual convention, Seattle, WA.

McMaster, K. L., Han, I., Chaffin, M. C., & Fuchs, D. (2013). Promoting teachers' use of scientifically-based instruction: A comparison of university versus district support. *Elementary School Journal, 113*(3), 303–330.

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention, 32,* 32–39.

Rafdal, B. H., McMaster, K. L., McConnell, S. R., Fuchs, D., & Fuchs, L. S. (2011). The effectiveness of kindergarten peer-assisted learning strategies for students with disabilities. *Exceptional Children, 77,* 299–316.

Saenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional Children, 71,* 231–247.

Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Saenz, L., Yen, L., Fuchs, L. S., & Compton, D. L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis, 30,* 368–388.

Part IV

Tier 2– Assessment, Problem Analysis, and Intervention

Assessment: Periodic Assessment to Monitor Progress

Benjamin Silberglitt, David Parker and Paul Muyskens

In the past 5–10 years, there has been a considerable increase in available research-based progress monitoring measures. Measures for mathematics and reading are widely accessible (e.g., Pearson 2012; University of Oregon Center on Teaching and Learning 2012; Interventioncentral.org), and national databases summarize the extent to which evidence supports using progress monitoring measures with students in various grades (e.g., National Center on Intensive Intervention 2013). The greater availability of progress monitoring measures allows for assessing learning rates for more students in more subjects, which strengthens the overall promise of response to intervention (RTI) because the assessment of learning rates is a cornerstone for successful RTI implementation (Batsche et al. 2005).

The focus of this chapter is on monitoring progress within tier 2 of an RTI system. How progress monitoring at tier 2 situated within a comprehensive RTI system is explained. Next, conceptual, technical, and practical considerations among progress monitoring approaches are described. Examples from various subjects and school levels are shared to illustrate promising practices. Finally, the chapter draws heavily from the robust research base on oral reading fluency (ORF), a specific assessment for reading that has been widely researched as a progress monitoring tool, to discuss issues related to implementing progress monitoring and directions for future research, as well as their implications for practice.

Context of Progress Monitoring at Tier 2

In most conceptualizations of RTI, a majority of students make adequate progress within the regular education system through the primary instruction provided by their teacher, referred to as tier 1 instruction. Benchmark testing or screening typically follows a tri-annual schedule that allows educators to identify those students not making adequate progress at key junctures within an academic school year (e.g., early in the fall; midway through the year). These are the students who might be provided interventions within tier 2 of the RTI framework, and for whom regular progress monitoring data are collected to inform various potential decisions about their performance during the intervention. Tier 2 of RTI is where the intensity of interventions is increased beyond what has historically been provided by regular education, and where the intensity of assessment is likewise increased beyond the tri-annual screening.

B. Silberglitt (✉) · P. Muyskens
TIES, Saint Paul, USA
e-mail: benjamin.silberglitt@ties.k12.mn.us

D. Parker
ServeMinnesota, Minneapolis, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_16

Increasing the intensity of assessment through progress monitoring is a critical component of the data-driven procedures that make up RTI, because it allows educators to assess progress on a frequent and consistent basis, thereby providing the necessary data to determine whether or not a student is responding to the corresponding increase in intensity of intervention. The progress monitoring data collected in tier 2 allow educators to make decisions including (a) if intervention has been successful at helping the student reach target performance and can be discontinued, (b) if the intervention is producing acceptable progress but should be continued, or (c) if the intervention is unsuccessful and a change is necessary (Stecker et al. 2008). The latter decision is what provides continuity across tiers 2 and 3 within an RTI context. Progress monitoring data in tier 2 help to clearly identify students who need the most intensive degree of support. Interventions for these students are delivered in tier 3, often referred to as the tertiary intervention stage.

Progress monitoring within tier 2 is typically conducted on a weekly basis, usually using curricular assessments from grade-level material. Having similar data for multiple students allows teachers to compare progress between students and to look at the effectiveness of an intervention that may be applied to many students through a standard protocol RTI intervention system (Christ et al. 2005). The progress data are often graphed, which allows teachers, parents, and students to easily see and interpret student progress in relation to the goal or to a benchmark standard when available. In many progress monitoring systems, the rate of student change or slope of improvement can be compared to the target or goal for improvement after 8–15 weeks of data collection. These data then form the basis for objective data-driven decision-making.

Conceptual Considerations for Progress Monitoring in Tier 2

Regardless of subject area or student level in school, progress monitoring involves collecting ongoing data on student performance. The data are typically produced using assessments that are either directly or indirectly based on curricular goals, which make progress monitoring a form of curriculum-based assessment (CBA; Tucker 1985). Researchers have identified two broad approaches to collecting progress monitoring data using CBA (Hintze et al. 2006; Fuchs and Deno 1991). The first, subskill mastery measurement (SMM), separates long-term curricular goals and content into short-term objectives and creates small domains of assessment items that correspond to each objective. An example of SMM might be teacher-based tests of common spelling patterns (e.g., CVCe) during the week (or weeks) in which they are taught. Data from SMM indicate when short-term objectives are met and instruction can proceed to the next objective along with a corresponding change in assessment content. The second approach, general outcome measurement (GOM), differs from SMM by including assessment content that focuses on broad curricular goals and is of constant difficulty across the school year (Fuchs & Deno 1991). Consistent difficulty permits the GOM approach to inform progress toward year-end expectations for performance. A common example of GOM is having the student read out loud from a grade-level text (i.e., oral reading fluency), which is considered a GOM approach because the act of reading out loud from grade-level text subsumes several other skills (e.g., letter- and word-level phonics skills, accurate and proficient reading of connected text) and should be of comparable difficulty.

Comparisons between the GOM and SMM approaches found comparative advantages and disadvantages of each. Repeated assessment using content of similar difficulty, as done using the GOM approach, provides an index of retention and generalization of previously learned skills, and it also avoids shifts in measurement difficulty between assessment administrations. In addition, research in reading from kindergarten to grade 6 demonstrated robust validity for GOM data across different curricula (Fuchs and Deno 1992), and the strengths of the GOM approach were not dependent on content sourced directly from the curriculum (Fuchs et al. 1990; Fuchs et al. 1992).

Despite the advantages and growing popularity of the GOM approach, two conceptual issues appear to limit its applicability across school levels and academic domains. The first is that research showing the GOM approaches that do not require content from the curriculum led to the development of curriculum-independent GOM approaches that are now widely used (e.g., University of Oregon, Teaching and Learning 2012; Pearson 2012). The often-confusing but very important point is that curriculum-independent GOM is not directly connected to any specific educational curriculum (e.g., Fuchs and Deno 1994; Fuchs et al. 1990), despite being considered synonymous with curriculum-based measurement (CBM; Deno 1985), which is a highly researched and popular form of CBA. For example, researchers and practitioners often use curriculum-independent reading passages for assessing ORF, and refer to them as progress monitoring procedures as CBM (e.g., Ardoin et al. 2013; Christ et al. 2012). In reality, the material from the school's reading curriculum is not used as part of the assessment. Such an issue may be largely semantic and of minimal conceptual import because of the correspondence between generic assessment material and the specific curricular expectations; however, it is nevertheless a source of confusion for educators less familiar with the purpose and benefits of progress monitoring.

The second conceptual issue concerns the potential for the GOM approach to monitor progress across academic subjects and school levels. This issue requires consideration of differences among educational constructs that affect assessment practices (Messick 1981). Not all educational constructs are the same. They can vary in terms of complexity (for which age or school level can be a sufficient proxy) and learning

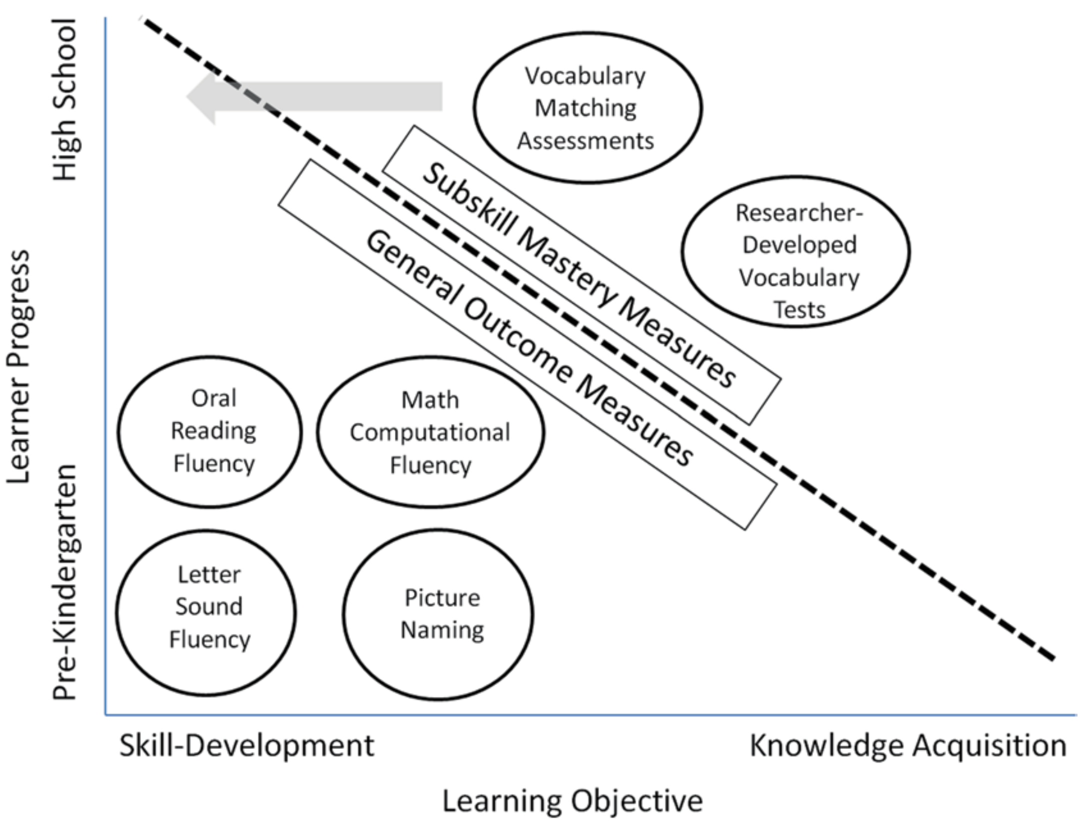


Fig. 1 A conceptual model of approaches to progress monitoring

objectives. For example, alphabetic knowledge is a key early literacy construct that consists of a constrained set of relatively simple learning expectations, including knowledge of letter names and sounds (Evans et al. 2006). Alternatively, reading comprehension is a much more complex task, requiring word-level decoding skills, fluent reading skills, content knowledge, the use of strategies, and application across literary styles (RAND 2002).

Another dimension on which educational constructs vary is the nature of the learning objective. Some constructs are mostly skill based while others are mostly knowledge based. For example, reading text accurately and proficiently is mostly skill based. On the other hand, vocabulary is a knowledge-based construct that requires receptive, expressive, and multi-level understanding of the meanings of words (Bornstein and Haynes 1998).

Differences in educational constructs likely affect the measurement of skill improvement over time, which is a core element of progress monitoring (Deno 1985; 2003). Some educational constructs (e.g., mathematical calculation; reading text fluently) may be more amenable to GOM progress assessment because skill improvements within the overall domain represent a sufficiently measureable improvement. Broader or more complex constructs may be less amenable to GOM progress assessment because actual increases in skill acquisition represent a change in improvement that is difficult to measure. Figure 1 represents a simple conceptual model of how different approaches to monitoring progress fall along two continua on which educational constructs can vary; school level is used as a proxy for complexity, because the developmental arc of complexity corresponds so closely with school level. The diagonal line is a rough indicator that GOM assessments tend to be more applicable with younger students or acquiring skills, whereas SMM assessments tend to be more applicable with older students or the acquisition of knowledge. Research indicates that each approach has, or has the potential for, the critical features necessary for instructionally useful assessment, including (a) repeatability with alternate forms of comparable difficulty, (b) defensible validity as an outcome indicator, and (c) provision of qualitative feedback on student performance (Fuchs and Deno 1994).

A more in-depth discussion of vocabulary can be used to illustrate the need for different approaches to progress monitoring with some educational constructs. Vocabulary is vitally important to overall reading performance (Anderson and Nagy 1992; National Reading Panel 2000), and a growing body of research has shown promising effects for vocabulary interventions within an RTI system (Loftus et al. 2010; Pullen et al. 2010), but absent from the extant research are measures and procedures for monitoring progress of individual students receiving vocabulary interventions. This dearth of progress monitoring limits the applicability of an RTI framework to vocabulary, due to an inability to measure the RTI (e.g., Batsche et al. 2005). One reason for the problem is that research estimated as many as 88,500 words exist in school-relevant English (Nagy and Anderson 1984), which means weekly increases in word knowledge would be potentially imperceptible relative to general curricular goals (even if divided proportionally between grade levels). In addition, unlike other skills for which GOM is effective, the construct of vocabulary consists of the acquisition of discrete knowledge (Stahl and Nagy 2006). Each word represents at least one unit of knowledge, rather than increased proficiency in a skill (e.g., decoding words with greater speed) or increased accuracy in a category of skill (e.g., single-digit mathematical fact computation). Given these facts, even pre–post analyses are suggested to be completed with researcher-developed measures instead of standardized assessments so as to detect measureable improvements (NRP 2000). Thus, at least pending future research, instructionally useful progress monitoring for vocabulary may need to employ SMM approaches in favor of GOM approaches (as shown by Fig. 1).

Vocabulary itself may not be currently amenable to GOM progress monitoring, but it has shown promise as a way to monitor progress in the content domains (e.g., Espin et al. 2005; Vannest et al. 2011). The procedures involve vocabulary matching tasks. The tasks do not use a GOM for progress monitoring, but because the content areas tend to have established curricular goals, they have been able to adapt some of the original principles of GOM for measuring growth toward global goals. For example, progress monitoring in social studies used a vocabulary matching task in which keywords from within the curriculum were identified and randomly ordered into a series of weekly assessments, which resulted in observable increases in social studies learning during progress monitoring (e.g., Espin et al. 2005). These results showed promise for applying the original GOM principle of creating progress monitoring assessments using curriculum-sourced content (Fuchs and Deno 1991).

Monitoring progress is essential to the successful implementation of tier 2 within an RTI framework. However, it seems that the particular approach taken to create a cycle of continuous, repeatable assessment of student growth will likely vary depending on the content domain being measured, as well as the interplay between learning objective and learner progress (often associated with school level) within that domain. These conceptual considerations are critical context within which technically sound strategies can be developed for monitoring progress at tier 2. Without this context in mind, the pursuit of a strategy (such as finding a GOM of vocabulary) may be a fruitless effort, in that technically adequate measures may not be possible. Choosing an appropriate strategy is a critical first step in establishing methods for monitoring progress. The next step is to ensure the selected method is technically sound.

Technical and Practical Considerations for Progress Monitoring at Tier 2

An important consideration for any data-driven decision is that the data themselves are technically appropriate for their intended use. For progress monitoring, the primary technical considerations consist of the reliability and validity of the data produced by the measures. An additional consideration is their sensitivity to growth, which is an important consideration given the importance of improvement within an RTI framework.

The reliability of progress monitoring measures indicates the degree to which the data are consistent, and is central to their ability to identify improvement. Both test-retest and inter-rater reliability have long been documented for several measures of progress monitoring including ORF (Tindal et al. 1983; Wayman et al. 2007) and CBM maze procedures (Marston 1989; Shin et al. 2000). In addition to reliability, progress monitoring also needs to produce valid data. Validity for measures used to assess progress is typically focused on predictive validity for some criterion outcome, or the ability to predict performance on a future assessment. Measures of ORF have been shown to have good predictive validity, which indicates the promise of these measures for students in the early grades (Andren 2010; Hintze and Silberglitt 2005; Wiley and Deno 2005). This predictive validity has been evidenced both within the same grade level (McGlinchey and Hixson 2004; Silberglitt et al. 2006), and across multiple grade levels (Baker et al. 2008; Hintze and Silberglitt 2005; Wanzek et al. 2010). GOM approaches to progress monitoring in other content domains, such as mathematics and writing, have also produced data that are sufficiently reliable and valid for progress monitoring (Foegen et al. 2007; McMaster and Espin 2007).

In addition to producing data that are reliable and that lead to valid decisions, progress monitoring measures should also be sensitive to change. Within tier 2, the focus is typically on short-term changes, with a common goal of determining whether or not an intervention is effective over the course of 8–10 weeks. GOM approaches to progress monitoring have been shown to be sensitive to change in student skill, and some GOM approaches are even sensitive to changes in measurement conditions (Christ and Hintze 2007). Factors affecting sensitivity to growth include variability in the probes being administered (Ardoin and Christ 2009), frequency of assessment, and overall change in performance over the course of time. This sensitivity heightens the need for assessment measures that have strong reliability and validity evidence, an issue that will be discussed later in this chapter.

For some educational constructs, the potential to assess change may need to be conceptualized differently. For example, the measurement of changes in overall vocabulary is exceptionally difficult given the breadth of the construct, but changes in acquisition of word meanings from 1 week to the next during an ongoing intervention might be important in terms of monitoring how well vocabulary intervention is working for students receiving it. To illustrate, if ten words were taught each week, and in a given week, nine of the words were understood on a weekly SMM, but the next week only four words were retained, then that is potentially useful information to those implementing the intervention. Additional research in this area is necessary to establish reliable and valid methods for measuring RTI where SMM is the appropriate viable measurement strategy.

In addition to technical considerations, progress monitoring in tier 2 of an RTI framework also requires consideration of practical issues. The wide application of these measures and the repeated nature of their administration necessitate attention to the efficiency and cost of the measures (Andren, 2010). Just as one would not use a measure that did not provide technically sufficient data, one would also not use a measure that took so long to administer that it supplanted significant instructional time or drained a building of resources. Achieving the balance between measures that are quick and efficient, and those that provide a broad range of useful and interesting instructional information is not always easy.

This balancing act between technical adequacy and efficiency is apparent by looking at the application of measures of ORF. Research on CBM for reading (CBM-)R and ORF has been going on for 35 years (Jenkins and Fuchs 2012), and at various times during its development, measures of ORF have varied in length, included comprehension questions and ratings of expression, used passages directly from instructional materials, or been parts of larger systems of reading assessment. The iterations of how ORF is measured are related to questions of efficiency and application and the expectations of the end user. An example of an effort to enhance the predictive validity of ORF was provided by a recent technical report by the Center for Teaching and Learning at the University of Oregon (2012) in which measures from the *dynamic indicators of basic early literacy skills next* (DIBELS Next; Good et al. 2013) were used to predict subsequent performance on a standardized group-administered reading test. The DIBELS Next measures, which included multiple measures of basic early literacy skills, were compared to measures of ORF. The DIBELS Next composite score included ORF, a retell score, and a measure of reading accuracy of words read. An examination of the data using a sequential regression procedure found that a fall administration of ORF explained approximately 40% of the variance in the spring score on the group reading test. Thus, there was not convincing evidence to recommend the administration of the items needed to calculate the composite score. Moreover, the authors stated that, “It has been difficult for us to find a measure that adds much to ORF’s prediction—even when all of the DIBELS measures are aggregated to form a composite” (p. 4, Center for Teaching and Learning at the University of Oregon 2012). Measures of ORF are very efficient and provide immediate feedback to the teacher, important elements in the provision of progress monitoring to tier 2 students.

ORF is perhaps the ideal GOM, in that it has the ability to function both as a screening and progress monitoring tool. As noted by Shapiro et al. (2006), “Given the expense and time required to administer norm-referenced achievement tests, CBM offers a potentially inexpensive way for districts to do large-scale screening” (p. 32). Including screening data in a progress monitoring graph can enhance the reliability of the slope through the inclusion of additional data points and provides a reference point for future performance on a subsequent high-stakes assessments, something that many measures of growth may not be able to do (Schatschneider et al. 2008). Put another way, the ability of ORF to predict future performance on high-stakes tests and to be sensitive enough to show responsiveness to intervention success provides the user the advantage of being able to graph progress not just in comparison to weekly rate of increase, but also to a criterion for future success (Deno 2003; National Center on Response to Intervention 2010).

ORF has thus become the de facto gold standard against which all other attempts to create GOMs are compared. This may seem an unfair standard to a researcher exploring a budding line of study on establishing a GOM of an educational construct other than reading, given the time and resources that have been invested in perfecting ORF. However, examining the history of the development of ORF, as well as the remaining concerns and continued research needed on this particular GOM for progress monitoring at tier 2, provides a useful road map for the development of GOM of other constructs. This chapter next briefly illustrates progress monitoring across educational constructs, and then specifically explores the continued challenges with ORF more deeply, as they are instructive both to ORF as well as to progress monitoring across constructs.

Illustrations of Tier 2 Progress Monitoring

As stated previously, progress monitoring helps educators make various data-driven decisions regarding intervention effectiveness. As a result, progress monitoring data contribute to beneficial student outcomes and are said to have high treatment validity (Cone 1989; Messick 1994). Although interventions themselves are the causal mechanism for improved outcomes, progress monitoring provides an empirical database with which to ensure interventions are effective for individual students, because even research-based interventions can produce idiosyncratic effects at the individual level (Eckert et al. 2002). Accordingly, empirical research shows that teachers who use data from progress monitoring make more decisions intended to improve instruction for students, thereby helping students achieve better outcomes (e.g., Fuchs et al. 1984; Stecker et al. 2005). What follows are two sample illustrations of how progress monitoring can be conducted in constructs other than reading fluency, and include a description of progress monitoring approaches in the secondary content areas.

Early Writing

Recent research has identified several GOM approaches to assessing early writing progress (e.g., Coker and Ritchey 2010; McMaster et al. 2009; McMaster et al. 2011). These measures involve timed administration of items that scaffold in idea generation, which is important because writing taxes idea generation of young students (Berninger and Amtmann 2003) and can potentially mask measurement of their skill growth. The available measures can be scored using various metrics (e.g., correctly spelled words, correct letter sequences) and have promising technical characteristics in terms of reliability and validity of single-point scores as well as for measuring growth (McMaster et al. 2011). Figure 2 shows a sample progress monitoring graph of a student who was identified as needing additional writing support. The baseline condition shows the student's performance prior to beginning intervention, which was markedly lower than peer comparisons with typical skills. Intervention 1 was initially effective, but ongoing data collection indicated that after week 6 the student stopped improving, which led to the implementation of intervention 2. Intervention 2 resulted in another considerable improvement in skill, and after 5 weeks showed a strong upward trend. In the current example, intervention 1 and intervention 2 were both research based, but the progress monitoring data were necessary to ensure that instruction transitioned to intervention 2 when the effects of intervention 1 were no longer beneficial for this individual student.

Monitoring Vocabulary Progress

Unlike early writing, progress monitoring for early vocabulary is less amenable to GOM approaches. The discrete nature of learning word meanings, the large number of words to potentially learn, and the complexity of word meanings all are challenges for the creation of GOM progress monitoring approaches. Moreover, early vocabulary does not have sufficiently established curricula from which to create GOM approaches to

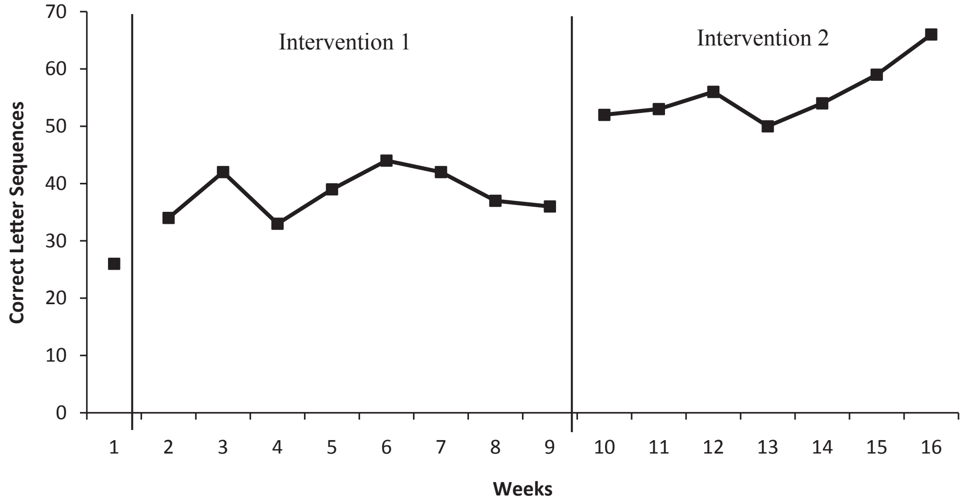


Fig. 2 Progress monitoring graph of a student needing additional writing support

progress monitoring. Instead, an SMM approach may be useful for monitoring progress during vocabulary interventions (e.g., Loftus et al. 2010). Figure 3 shows a bar graph depicting 8 weeks of vocabulary progress monitoring for two students, one higher performing and one lower performing. Both students were in the same research-based intervention (Beck and McKeown 2007), but student 1 consistently learned nine or ten out of the ten target words taught each week. Student 2 was much less consistent, learning as few as five of the ten targeted in a week. These data would be useful for providing additional instruction and support for student 2.

Monitoring Progress in the Content Areas

Measures for monitoring progress in the content areas have typically used SMM approaches. This is likely because learning within the content

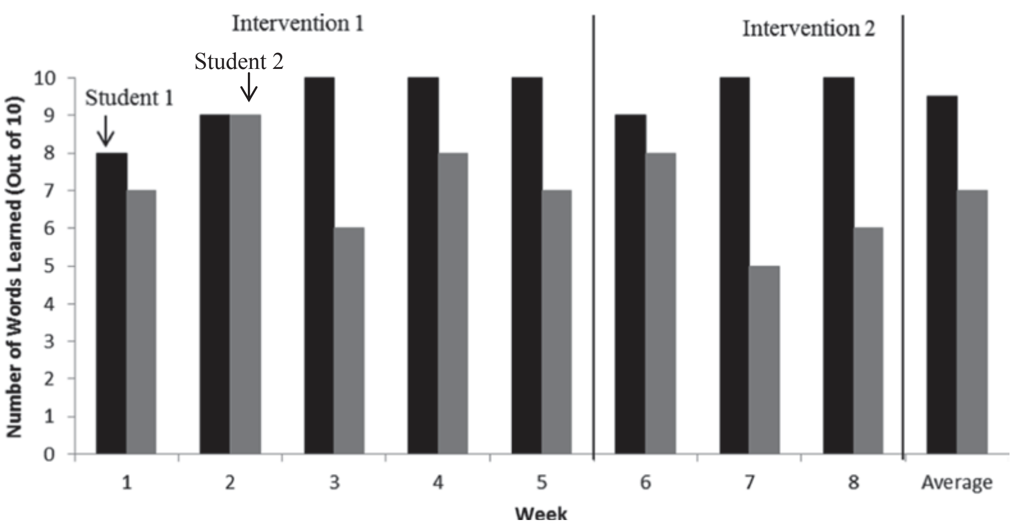


Fig. 3 Progress monitoring in vocabulary for two students

areas is so compartmentalized. A science unit on biomes is dramatically different from a science unit on cells, and a social studies unit on the judicial branch of government is different than a unit on cities. Despite the challenges in creating measures that have GOM characteristics, researchers have developed and tested approaches to progress monitoring in the content areas that share some characteristics of GOM approaches (this is represented by the grey arrow in Fig. 1). These procedures involve vocabulary matching tasks.

In one approach for social studies, researchers randomly sampled words and definitions from the local social studies curriculum, organized the words and definitions (plus two distractors) into 11 different probes, and collected weekly progress monitoring to determine if the approach was sensitive to growth (Espin et al. 2005). Results indicated the measure was sensitive to growth. In science, a different research team used a more comprehensive review of science curricula to develop six sets of 20-item progress monitoring probes, with results that showed the approach was sensitive to growth in student science knowledge (Vannest et al. 2011).

Key Issues, Implications for Practice, and Directions for Research

The possibilities for monitoring progress in tier 2 of an RTI framework are extensive. What began as an isolated attempt to better understand elementary age students' growth in reading (Deno and Mirkin 1977) has been extended to early childhood and high school, and to a broad range of educational constructs. The expansion of this measurement approach is no doubt because of its inherent value for instructional decision-making. Having access to weekly or monthly assessment data to track student growth over short increments of time provides excellent and timely information, so that informed decisions about tier 2 interventions can be made while the interventions are ongoing, before months of instructional time have passed. This is certainly an exciting time for practitioners interested in growing what is known as a highly effective model in elementary age reading, and replicating the RTI model across ages and content areas.

Although there has been recent improvement in progress monitoring technology in reading and other content areas, significant measurement issues remain with the assessments commonly used to monitor progress, and it is important for the practitioner to consider these issues when making instructional decisions. The following section examines these considerations, providing both direction for future research and discussion of the practical implications of the current state of available assessments. This section examines four key issues that are necessary for understanding growth and providing adequate guidance for practitioners engaged in progress monitoring in a tier 2 setting: reliability and validity of change, sensitivity of change, linearity of change, and standards and expectations for change.

Of all areas of progress monitoring, ORF has the longest history (Deno 1985; Marston 1989) and the widest use. These four key issues will be presented almost exclusively around the topic of ORF. However, due to the importance of measuring student response in all tier 2 interventions for RTI (Batsche et al. 2005), the issues and guidance in this section are considered illustrative for all progress monitoring measures.

Reliability and Validity of Change

A key element of decision-making within a progress monitoring framework is the ability to establish an individual student's rate of progress, and draw comparisons between that rate of progress and an expected rate of progress. In one popular approach to a progress monitoring decision-making framework, the student's trend line is compared to the aim line, and decisions about instruction are made according to this comparison. If the trend line is below the aim line, ostensibly some change is needed, either in the intensity of the intervention or the nature of the intervention altogether. The other approach, the "data point approach," examines the number of consecutive data points above the aim line in order to make a decision about the success of the intervention

(National Center on Response to Intervention 2010). In either approach, it is the rate of growth or performance at multiple time points, rather than performance at a single-time point, that becomes the unit of analysis. While there is some question as to whether classical test theory is sufficient to examine growth as a unit of analysis, it still is helpful for informing our thinking conceptually about the reliability of our decision-making process. Considering classical test theory for a moment, the student's trend line is their "observed score," which is an estimate of their "true score." In theory, the observed score/true score distinction can be ignored when it comes to the aim line, because that is in effect a "true score" needed to demonstrate adequate performance (a criterion). However, other issues such as linearity and how to determine standards and expectations for change are especially relevant to the aim line and will be discussed later.

If the student's trend line is their observed score, then we must consider what factors would prevent this observed slope of growth from providing an exact estimate of their true slope of growth, the degree to which we can expect the observed score to vary around the true score, and the degree to which we can expect the observed score to be measuring what we intended. Understanding these issues will also help us to answer key questions that face the practitioner: How many data points do I need to gather, before I can be confident in the estimate of the student's slope? How do I establish a discrepancy between student progress and the aim line? How meaningful is the student's slope of growth, in predicting outcomes?

Research on Reliability and Validity of Change Christ and Silberglitt (2007) provided some context to the difficulties associated with establishing a reasonable degree of error with growth estimates by first examining the error associated with a point-in-time estimate. This study explored the standard error of measurement (SEM) across thousands of students' benchmark ORF assessments. The SEM was estimated using classical test theory to be between 5 and 15, depending on the estimates used for reliability, with a median SEM of 10. This study has two important practical implications for progress monitoring. First, an SEM of 10 means that a 68 % confidence interval around the students observed score would be plus or minus 10, and a 95 % confidence interval would be plus or minus 20. Second, this study was based on a median score of three passages, as is typically used in benchmark (tier 1) assessment, rather than a single passage score as is typically used in progress monitoring (tier 2 or 3) assessment. While no research has specifically examined the SEM of a single passage score, even assuming generously that 10 should be used, this has significant implications for the reliability of the growth estimates generated by progress monitoring data. If each point in time has such a wide confidence interval, then how does this impact the accuracy of developing a trend across multiple points in time, and what does this mean for using the spring benchmark target as the goal for an aim line, as is often done in tier 2 progress monitoring?

Fortunately, Christ and Silberglitt (2007) only looked at assessment at a point in time, and one would expect that the process of collecting multiple weekly data points would reduce the error associated with our estimates. While SEM is informative and should empower educators with an appropriate level of caution in interpreting results of a single ORF assessment, the most important error estimate associated with progress monitoring is the *standard error of the slope* (SE*b*), because the slope is the unit of analysis when monitoring progress. Christ (2006) provides an initial exploration into SE*b* estimates for ORF progress monitoring, using simulations to establish estimates ranging from 0.08 to 16.54 words per minute per week, depending on factors such as the *standard error of the estimate* (SEE) and the number of data points gathered. When looking at typical conditions in a practical setting, such as 6–8 weekly data points and a moderate SEE, the SE*b* estimates ranged from 0.87 to 2.01 words per minute per week, which should cause some hesitation, because the rule of thumb in practice has tended to be that 6–8 weeks of data are sufficient to establish the effectiveness of an intervention (Ardoin et al. 2013). In this condition, even generously assuming an SE*b* of 1.0, a 95 % confidence band around a student's slope would be plus or minus 2 words per minute per week. Given that another rule of thumb is that 2.0 words per minute per week is an ambitious yet attainable aim line (Fuchs et al. 1993), and that in the case of tier 2 many students will have aim lines with slopes less steep than 2.0, this line of research is certainly worthy of further exploration.

Further research by Christ et al. (2013) makes suggestions of 8 weeks of data for most decisions and 12–14 weeks for high-stakes decisions. However, research may still be insufficient for setting any hard and fast rule about the number of data points needed, and expert judgment is likely to continue to be an essential part of the process (Shapiro 2013).

Not surprisingly, research on the validity of slope has yielded similarly mixed results. Stage and Jacobsen (2001) performed one of the first studies exploring the validity of the slope of a student's growth rate. Specifically, they examined the predictive validity of slope on ORF in explaining variability on state-mandated assessments of reading. While they did find that slope was a significant predictor, they also found that the level of performance accounted for much more of the variance in performance on the criterion measure. Interestingly, they also found a moderate positive correlation between slope of growth and level of performance, which has implications for establishing expectations of growth that will be discussed later.

Clarke et al (2008) examined the predictive validity of slope with the *tests of early numeracy* (TEN; Clarke and Shinn 2004), and found that for all measures except quantity discrimination (QD), slope did not account for a significant amount of the variance on the outcome measure. While QD did explain some of the variance, it was also a much weaker predictor than level of performance.

Implications for Practice These outstanding questions of reliability and validity of growth measurement have significant implications for practitioners using GOM to monitor progress within a tier 2 setting. Certainly, the questions of SEM and SE*b* have implications for the process of decision-making within a progress monitoring graph. Figure 4 shows an example of the impact of error bands around the slope estimate of a hypothetical student, using an SE*b* of 1.0. In

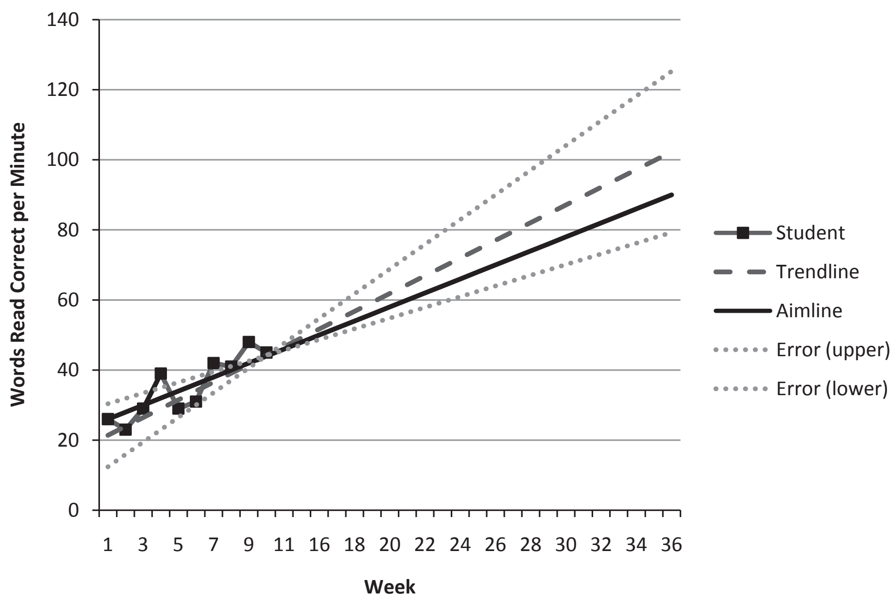


Fig. 4 Progress monitoring graph with error bands to reflect SEb of 1.0

this situation, without the error bands included, it is likely that most practitioners would have declared the intervention a success, and considered phasing it out. However, with the error bands the confidence in this decision is reduced.

The complexities of incorporating the error associated with slope estimates most likely require on-the-fly recalculations of error bands around trend lines, as each new datum is added to the graph. This likely requires the involvement of technology, yet to date it does not appear that software exists to support using SE*b* appropriately within a progress monitoring framework. In the meantime, practitioners must at least consider the error associated with these assessments, and encourage educators to triangulate sources of data and gather sufficient amounts of information to be very confident in their decisions.

Additionally, the question of whether slope accounts for a meaningful proportion of the variance in outcomes, over and above what is accounted for by level of performance, has implications for a dual-discrepancy model that cannot be understated. The concept of a dual-discrepancy approach within an RTI framework assumes that a student must be discrepant on both level and slope to be considered nonresponsive to an intervention (Fuchs and Fuchs 1998). However, if slope does not add value to the overall utility of the data for predicting student success, then why include it in the model? Despite the statistical argument for removing data that does not improve overall prediction, using both screening and progress monitoring data in practical settings does impact decision-making (Shapiro et al. 2012). While incorporating both screening and progress monitoring data certainly has great practical value for decision-making, better guidance is still needed as to how best to apply decision rules to these data.

Directions for Future Research Growth validity is a necessary direction for future research. In laying out a framework for developing new GOMs, Fuchs (2004) describes three stages of analysis: Stage I is to establish technical adequacy at a point in time; stage II is to establish technical adequacy of slope, and stage III is to establish instructional utility. Similar to the notion from classical test theory, that a measure cannot have validity if it does not possess adequate reliability, Fuchs (2004) argues that a general outcome measure cannot have instructional utility if it does not have reliability and validity both of level of performance and of slope. Practitioners need answers to questions about how wide of an error band to consider around a trend line, and whether a growth rate is necessary for establishing nonresponse to intervention or whether persistent discrepancies on level of performance are sufficient.

Sensitivity of Change

When monitoring student progress in a tier 2 setting, not only must the growth demonstrated be reliable and valid, it is essential that the general outcome measure be sensitive to change. That is, the measure itself must be able to detect the effects of differences in the instructional environment, and show changes in slope accordingly (Fuchs 2004). In a progress monitoring graph, the ability to detect a change in trend across a phase change is critical to decision-making (Baer et al. 1968). If the measure is not adequately sensitive, then slopes will appear similar across these phase changes, despite differing impacts of the intervention phases on a student's actual rate of growth.

Research on Sensitivity of Change An example where sensitivity can be a factor is when the general outcome measure demonstrates a floor effect. If the student being monitored is not able to demonstrate the skill using the measurement framework, then a different measurement is needed, in order to provide more dynamic information about the student's growth. An analogy from basketball is the child who cannot yet make a free throw. Until the student can make a free throw, using the number of free throws made successfully in 1 min as a measure of growth would fail to adequately capture the growth that the child may otherwise be making in their ability to shoot the basketball. Similarly, measuring a child's reading from connected text when they are still working on phonemic awareness will fail to capture the growth they make in pre-literacy skills. One might even argue that the GOM is not interval-level data, in the sense that the distance between 0 and 1 is greater than the distance between any two other adjacent numbers on the scale.

The issue of a floor effect is especially prevalent in the research on GOMs in early literacy and numeracy (Hojnoski et al. 2009; Missall et al. 2007). Often, these GOMs will produce mean scores on a norm sample that are smaller than two, and sometimes even one times the standard deviation, which makes it difficult to establish discrepancy based on the level of performance. Further research is needed on the impact of floor effects on growth calculations.

Implications for Practice The most apparent implication of a lack of sensitivity is that an intervention may have an effect that is under realized. If the assessment is not sensitive, a student may be believed to be dually discrepant, when in fact they were responding to the intervention but the assessment was not detecting the response. Especially in early skill development, where floor effects are more likely, it is essential that the practitioner select an assessment appropriate to the level of skill being instructed. This is the fundamental rationale for the practice of *survey level assessment* (Shinn 1989). If a useful GOM appropriate to the level of skill being instructed is not available, then the advice to the practitioner is to look to other sources of data to establish whether progress is being made, before making a change to instruction based solely on a flat trend on a progress monitoring graph.

Another implication of a lack of sensitivity is on the frequency of the assessments themselves. Tier 2 progress monitoring approaches typically suggest that intensity of assessment is increased along with intensity of intervention for students who are nonresponsive to intervention (Fuchs and Fuchs 2006). However, increasing the frequency of an assessment that is not sensitive to growth is not warranted. For example, the Northwest Evaluation Association produces a popular computer-adaptive assessment of reading and mathematics, known as the *measures of academic progress* (MAP; Northwest Evaluation Association 2013). The MAP is often given in the fall, winter, and spring of a school year, to provide some idea of growth over the course of the year. However, the average fall-to-winter growth of a 5th grader at the 50th percentile on the MAP reading is 3.4 points, which is about the same as a typical standard error of measurement on the assessment (Northwest Evaluation Association 2011). Thus, growth of an average student from fall to winter is likely already indistinguishable from error. By increasing the frequency of this assessment, the problem would only be exacerbated.

Directions for Future Research A key factor impacting the lack of research on sensitivity to change is the fact that, to date, the large majority of studies on progress monitoring have tended to be simulations (Christ et al. 2013). Simulation studies are useful in their ability to generate large volumes of data that would otherwise be expensive to collect. However, they are limited in their ability to produce useful guidance for practitioners. Future research should attempt to establish the sensitivity of GOMs in the presence of interventions with known effectiveness, on reasonably large samples of actual students. This research could also provide better guidance for decision-making on when to make a change in the intervention.

Linearity of Change

Progress monitoring graphs traditionally include a linear trend line, perhaps because they are easier to interpret and perhaps because of the roots of progress monitoring are using a paper and pencil graph and calculating a trend line by hand. However, there is strong evidence that the slope of growth on many general outcome measures is typically nonlinear.

Research on Linearity of Change Silberglitt (2009) examined thousands of fall, winter, and spring ORF benchmark scores using both a piecewise and linear growth model within a linear mixed model approach. Across grades 2–5, and across regular and special education students, the piecewise model was a better tool for estimating growth across the year than the linear growth model because the piecewise model allowed for a change in slope at the winter time point, in this case with growth rates declining from winter to spring. The data from the Silberglitt (2009) study were reanalyzed using a latent growth model, with similar results (Christ et al. 2010).

Other studies have also found nonlinear within-year growth on both ORF and Maze, despite that the patterns of growth (i.e., whether fall–winter growth rates are steeper than winter–spring) were inconsistent both within and across studies (Ardoin and Christ 2008; Graney et al. 2009). Thus, it is not clear if nonlinearity is a function of the instructional environment or a characteristic of the assessments themselves (Graney et al. 2009). Certainly, it is possible that the measures are in fact responding to influences such as a curriculum where a significant period of time is focused on test preparation over traditional instructional methods, or where the implementation of the curriculum is less rigorous at the end of the year, for example.

Implications for Practice From a practical standpoint, nonlinearity of growth needs to be considered in both the calculation of the aim line and the trend. Incorrectly assuming linearity across the school year can have significant consequences for students. Take the case where fall to winter growth outpaces winter to spring. In this setting, students who are monitored in the beginning of the year will tend to have steeper slopes, and thus be more likely to be considered responsive to intervention than students monitored at the end of the year. Further, if a trend line is extrapolated for a student in the beginning of the year, it may appear falsely that they are on-track to meet spring targets.

Figure 5 presents an example of a second-grade student, who begins the year reading 18 words correctly per minute, which is below the tier 1 target of 43. Although these data are simulated for a fictitious student, they are based on the average tier 2 progress monitoring case in a region of Minnesota in which thousands of students are participating in a federally funded tier 2 intervention program. The average was modeled using a linear mixed model, which established a nonlinear trend with a negative quadratic component. The average case was used to show how widespread this type of misinterpretation of

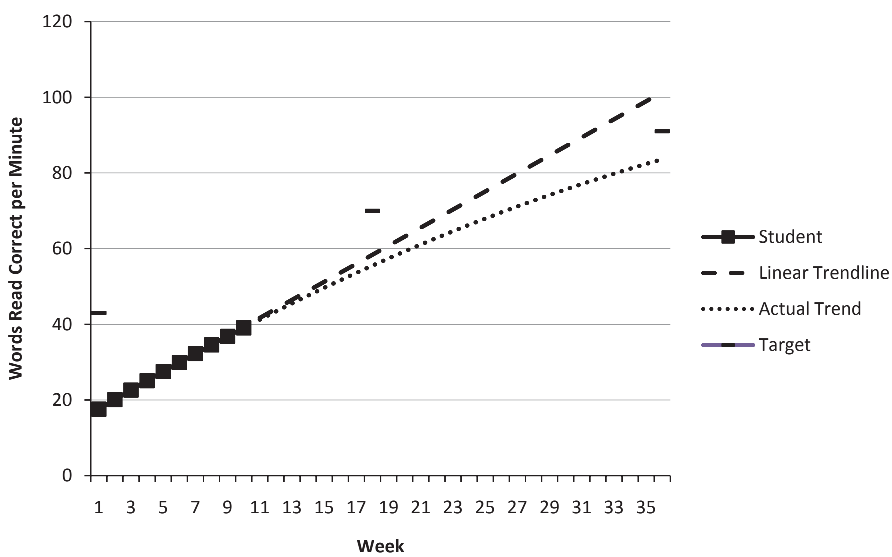


Fig. 5 The inaccuracy of setting a linear trend line when growth is nonlinear

a progress monitoring graph likely is. The student's progress is monitored for 10 weeks, and grows according to the model. A linear trend line shows how significant the discrepancy is between the predicted growth and actual growth of the student, where their actual growth continues according to the district model. The linear prediction has the student on track to far exceed the spring benchmark target, a situation where most practitioners would start exploring phasing out the intervention, while the more accurate nonlinear model shows that the student should probably be getting more intensive intervention.

The school personnel involved in the federally funded RTI initiative decided to make their exit criteria more stringent because of the nonlinearity of the data. Thus, a student who may have previously been exited from intervention if they demonstrated a number of consecutive data points above the aim line now needs to demonstrate both performance above the aim line with an increased number of data points and performance above the upcoming season's benchmark target score. Modifying decision rules in this manner seems warranted, given that the risks of removing valuable intervention services for a student prematurely are greater than the risks of providing too much intervention.

Directions for Future Research Future research should explore whether nonlinearity of growth can be reasonably predicted by the instructional environment, or whether it occurs across differing levels of instructional intensity. If the latter is true, then further research is needed to provide better guidance to practitioners. Both aim lines and trend lines would need to be modeled using nonlinear regression models, in order to provide accurate predictions of outcomes for students. Large-scale data collection efforts would greatly assist in establishing typical levels of nonlinear growth across general outcome measures, so that models of individual student growth can be based on reasonable expectations. The widespread use of computer software to calculate trend lines means that both the data collection effort and the ability to calculate complex nonlinear trends are increasingly feasible in a practical setting. While practitioners wait for this needed research, however, continuing to monitor student progress well after the phasing out of an intervention seems warranted.

Standards and Expectations of Change

Establishing appropriate expectations for change during progress monitoring is a key element of the process of decision-making. In a dual-discrepancy model, it is as important to accurately estimate the student's current state as it is to accurately estimate from what the child is discrepant. For slope, this expectation is typically established via the aim line. This aim line is often set based on what is typical growth, or alternatively what growth is needed to "catch up" to typical.

Research on Standards and Expectations of Change Research suggests that a typical level of growth may be difficult to establish. As stated earlier, Stage and Jacobsen (2001) found a correlation between level of performance and slope on ORF, indicating that growth rates were not equivalent at all levels of performance. Moreover, Silberglitt and Hintze (2007) found that there were wide variations in average growth on benchmark assessments across deciles based on fall performance. Specifically, lower and higher deciles tended to show lower rates of growth than deciles near the median, with more pronounced differences at earlier grade levels.

Implications for Practice This variability in typical levels of growth has direct implications for practitioners attempting to implement a dual-discrepancy model. For instance, should the growth rate of a student in a lower decile on level of performance be compared to the growth rate of students at the median, or to the growth rate of students at their decile? Given the variability of growth rates across deciles, the latter is not practically feasible. Additionally, there is a serious concern with establishing lower growth standards for lower-performing students, simply because that is what is typical. However, a comparison to growth for higher-performing students may not be relevant or useful.

One consideration for practitioners is whether a normative approach to establishing dual discrepancy is appropriate. In fact, concerns regarding the practice of making high-stakes educational decisions using arbitrary and mercurial rules based on norms are what led us to the dual-discrepancy model in the first place, as an alternative to the IQ-achievement discrepancy model. A practical strategy that can have immediate value is to examine discrepancy of growth (as well as discrepancy of level) as discrepancy from expectation rather than discrepancy from peers. When employing a system of criterion-referenced target scores on periodic screening assessments, a meaningful, formative guideline for both level of performance and rate of progress is obtained. The rate of change of these target scores from fall to spring becomes a stable benchmark of 1 year's growth in 1 year's time, and discrepancy of growth is established as whether the student is meeting that benchmark, or perhaps demonstrating growth that will catch them up to target in a reasonable time frame.

Directions for Future Research Research on establishing discrepancy within a criterion-referenced model would have direct and immediate practical value. Increasingly, the target scores on GOMs are being established based on their prediction of state-mandated assessments (Reschly et al. 2009). What is not known is the degree to which this presents variability across states, or how this will be impacted as states move toward new collaborative state assessments such as those being developed by the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC). Additionally, while it is relatively easy to establish an aim line that ends at spring target, research is needed to help inform practitioners as to when that aim line is appropriate, and when and how aim lines may be established for significantly underperforming students, who may need more than 1 year to catch up to target.

Conclusions

Progress monitoring is a critical aspect of tier 2 decision-making within an RTI framework. As this framework continues to gain popularity, practitioners are looking for ways to stretch the boundaries of what is known in terms of how to monitor progress in ways that are both technically sound and instructionally useful. From its roots in early elementary literacy assessment, the concept of progress monitoring is now being applied across early childhood through high school, and on an increasingly broad range of educational constructs.

Certainly, a context is needed for this expansion, as practitioners and researchers wrestle with how to develop assessments that can fit this framework. This push for assessment information is an indication of the value of this decision-making framework, and puts educators in the right mindset, in terms of developing assessments with a meaningful purpose in mind, a priori. However, it also may be true that the boundaries of what GOMs can provide are reached. In some ways, ORF was both the best of GOMs and the worst of GOMs, in that no other GOM to date can match its simplicity and elegance for informing instructional decisions. As practitioners invariably drift to SMM assessment for some grade levels and educational constructs, it is hoped that providing some context around these decisions will help practitioners consider which approach may be best.

Even as the use of GOMs such as ORF becomes more widespread, there remain a significant number of questions to be answered regarding the implementation of ORF assessment in an RTI framework. Continuing to pursue the answers to these questions will inform practice and inform the development of other GOM assessments designed to replicate ORF's success in other grade levels and educational constructs. Table 1 summarizes some of the key issues that impact decision-making using ORF progress monitoring data in tier 2, provides directions for future research to address these issues, and provides practical considerations for how to deal with these issues while waiting for research to provide better information.

Regardless of the road ahead, the impact and value of collecting frequent assessment data to inform progress is without question. Using these data to guide our decision-making is a quantum leap over previous instructional decision-making frameworks. Collecting these data and helping educators to understand and interpret them leads to better instructional decisions (Shapiro 2013). In fact, the act of monitoring progress itself can even positively impact student performance (Fuchs et al. 1984). However, it is our responsibility as educators to understand the limitations of the available set of progress monitoring assessments. While the RTI framework increases in popularity, so too do the risks associated with bringing this model to scale. Maintaining appropriate and consistent practice is made even more difficult when the research on many key questions impacting implementation is still evolving. The temptation is strong to simply create a rule that will give educators hard and fast answers to these questions, and will guarantee consistency at scale. Moreover, rules such as 3 data points above the aim line and 10 weeks before making

Table 1 Key issues in the use of oral reading fluency (ORF) for progress monitoring in Tier 2

	Reliability and validity of change	Sensitivity of change	Linearity of change	Standards/expectations for change
Key issues	High standard error of measurement	Concerns when floor effects may be present	Growth may be nonlinear	Norms for growth are unknown
	High standard error of slope		When growth is nonlinear and in what direction is unknown	Growth norms may vary with level of performance
	Low evidence for validity of slope			
Future research	Further stage II research needed	Research beyond simulation studies	Explore nonlinearity as a function of instructional environment	Examine dual discrepancy within a criterion-referenced model
		Test sensitivity against known effective interventions		Examine variability in criterion across states (and common core)
				Establish guidance for “catch-up” aim lines for students significantly below target
Implications for practice	Consider potential error of slope when making decisions	Consider whether GOM is appropriate to level of skill	Make exit criteria more stringent	Use target scores to establish criterion for growth in a dual-discrepancy model
	Incorporate other sources of data in decision-making	Do not increase frequency of assessments beyond their sensitivity	Continue monitoring well after exit	
	Gather sufficient information to have high confidence in decisions			

GOM general outcome measurement

a trend line comparison are often fairly arbitrary and have no research supporting them.

It may be helpful to remind readers that arbitrary decision rules around performance on achievement and IQ tests are partially what led to the downfall of the IQ–achievement discrepancy model. For example, regardless of the number of tweaks that are made to yield tiny incremental improvement in the SEM of ORF, or the continued research on decision rules, it is doubtful that this assessment can ever be turned into a laser that can singularly and definitively answer all of the questions that need to be answered within an educational decision-making framework like RTI (S. L. Deno, personal communication, November 18, 2013).

We should continue to press forward with establishing and implementing effective methods of periodic assessment to monitor progress within tier 2. At the same time,one should remember the limitations of these assessments, and encourage decision-makers to consider multiple sources of data as well as expert judgment, especially when making high-stakes decisions in educational settings.

References

Anderson, R. C., & Nagy, W. E. (1992). The vocabulary conundrum. *American Educator, 16*(4), 14–18, 44–47.

Andren, K. J. (2010). An analysis of the concurrent and predictive validity of curriculum based measures (CBM), the measures of academic progress (MAP), and the New England common assessment program (NECAP) for reading (Doctoral dissertation, University of Southern Maine, 2010). Proquest, Umi Dissertation Publishing, 2011.

Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review, 37*(1), 109–125.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, Aimsweb, and an experimental passage set. *School Psychology Review, 38*(2), 266–283.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Baker, S. K., Smolkowski, K., Katz, R., Hank, F., Seeley, J. R., Kame'enui, E. J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37,* 18–37.

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., … Tilly, W. D. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria: National Association of State Directors of Special Education.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251–271.

Berninger, V., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345–363). New York: Guilford.

Bornstein, M. H., & Haynes, M. O. (1998). Vocabulary competence in early childhood: Measurement, latent, construct, and predictive validity. *Child Development, 69,* 654–671.

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*(1), 128–133.

Christ, T. J., & Hintze, J. M. (2007). Psychometric considerations when evaluating response to intervention. In S. Jimerson, M. Burns, & A. VanDerHeyden (Eds.), Handbook of response to intervention: The science and practice of assessment and intervention. New York: Springer.

Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measurement of oral reading fluency. *School Psychology Review, 36*(1), 130–146.

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *NASP Communique*, *34*(3), 1, 6–8.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum based measurement of oral reading (CBM-R): An evaluation of linear expectations for growth. *School Psychology Review, 39*(3), 447–462.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. R. (2012). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology, 51,* 19–57.

Clarke, B., & Shinn, M. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33,* 234–248.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29*(1), 46–57.

Coker, D. L., & Ritchey, K., D. (2010). Curriculum based measurement of writing in kindergarten and first grade: An investigation of production and qualitative scores. *Exceptional Children, 76,* 175–193.

Cone, J. D. (1989). Is there utility for treatment utility? *American Psychologist, 44,* 1241–1242

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37,* 184–192

Deno, S. L., & Mirkin, P. K. (1977) *Data-based program modification: A manual.* Reston: Council for Exceptional Children.

Eckert, T. L., Ardoin, S. P., Daly, E. J., & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35*(3), 271–281.

Espin, C. A., Shin, J., & Busch, T. W. (2005). Curriculum-based measurement in content areas: Vocabulary matching as an indicator of progress in social studies learning. *Journal of Learning Disabilities, 38,* 353–363.

Evans, M.A., Bell, M., Shaw, D., Moretti, S., & Page, J. (2006). Letter names, letter sounds and phonological awareness: An examination of kindergarten children across letters and of letters across children. *Reading and Writing, 19,* 959–989. doi:10.1007/s11145-006-9026-x.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics a review of the literature. *The Journal of Special Education*, *41*(2), 121–139.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33,* 188–192.

Fuchs, L. S., & Deno, S. L., (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–500.

Fuchs, L. S., & Deno, S. L., (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232–243.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61,* 15–24.

Fuchs, L., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13,* 204–219.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on student achievement, pedagogy, and student awareness of learning. *American Educational Research Journal, 21,* 449–460.

Fuchs, L. S., Allinder, R. M., & Fuchs, D. (1990). *Drawing measurement samples from the curriculum. Effects of instructional planning and student achievement. Unpublished manuscript. based reading practices for response to intervention.* Baltimore: Paul H. Brookes Publishing Co.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert based system consultation within curriculum-based measurement using a reading maze task. *Exceptional Children, 58,* 436–450.

Fuchs, L.S., Fuchs, D., Hamlett, C.L, Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27–48.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2013). *DIBELS Next Technical Manual*. Eugene: Dynamic Measurement Group.

Graney, S. B., Missall, K. N., MaRtInez, R. S., & Bergstrom, M. (2009). A preliminary investigation of within-year growth patterns in reading and mathematics curriculum-based measures. *Journal of School Psychology, 47,* 121–142.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high stakes testing. *School Psychology Review, 34,* 372–386.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*, 45–56.

Hojnoski, R. L., Silberglitt, B., & Floyd, R. G. (2009). Sensitivity to growth over time of the preschool numeracy indicators with a sample of preschoolers in head start. *School Psychology Review, 38,* 402–418.

Jenkins, J. R., & Fuchs, L. S. (2012). Curriculum-based measurement: The paradigm, history and legacy. In C. Espin, K. McMaster, S. Rose, & M. Wayman (Eds.), A measure of success: The influence of curriculum-based measurement on education. University of MN Press, Minneapolis.

Loftus, S., Coyne, M. D., McCoach, D. B., Zipoli, R., Kapp, S., & Pullen, P. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at-risk for language and literacy difficulties. *Learning Disabilities Research & Practice, 25*(3), 124–136.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York: Guilford.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 193–203.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing a literature review. *The Journal of Special Education, 41*(2), 68–84.

McMaster, K. L., Du, X., & Petursdottir, A. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities, 42,* 41–60.

McMaster, K. L., Du, X., Yeo, S., Deno, S. L., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children, 77,* 185–206.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin, 89,* 575–588.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2)*,* 13–23.

Missall, K. N., Reschly, A., Betts, J., McConnell, S., Heistad, D., Pickart, M., et al. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review, 36,* 433–452.

Nagy, W., E., & Anderson, R., C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304–330.

National Center on Response to Intervention. (2010). Brief #3. Common progress monitoring graph omissions: Making instructional decisions. Washington, DC: US Department of Education, Office of Special Education Programs. http://www.RtI4success.org. Accessed 14 May 2010.

National Center on Response to Intervention. (2010). Progress monitoring tools Chart: Reading and Math. http://www.RtI4success.org/chart/progressMonitoring/PMToolsChart_04-20-10a.pdf. Accessed 14 May 2010.

National Center on Intensive Intervention. (2013). Academic progress monitoring tool. http://www.intensiveintervention.org/chart/progress-monitoring. Accessed 20 Dec 2013.

National Reading Panel. (2000). *Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Bethesda: National Institute of Child Health and Human Development.

Northwest Evaluation Association. (2011). *RIT Scale Norms Study*. Portland: NWEA.

Northwest Evaluation Association. (2013). *Measures of academic progress*. Portland: NWEA.

Pearson. (2012). AIMSweb progress monitoring and improvement system. http://www.aimsweb.com/. Accessed 29 Dec 2012.

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice, 25,* 110–122.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. Santa Monica: RAND.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47,* 427–469.

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences, 18*(3), 308–315.

Shapiro, E. S. (2013). Commentary on progress monitoring with CBM-R and decision making: Problems found and looking for solutions. *Journal of School Psychology, 51*(1), 59–66.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19–35.

Shapiro, E. S., Hilt-Panahon, A., Gischlar, K. L., Devlin, K., Leichman, E., & Bwles, S. (2012). An analysis of consistency between team decisions and reading assessment data within an RTI Model. *Remedial and Special Education, 33,* 335–347.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of maze task for curriculum-based measurement of reading growth. *Journal of Special Education, 34,* 164–173.

Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 90–129). New York: Guilford.

Silberglitt, B. (2009). *Curriculum Based Measurement of Oral Reading (CBM-R): An Evaluation of Growth Rates and Season Effects among Students Served in General and Special Education* (Unpublished Manuscript).

Silberglitt, B., & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children, 74*(1), 71–84.

Silberglitt, B., Burns, M. K., Madyn, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43,* 527–536.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30,* 407–419.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. Mahwah: Erlbaum

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795–819.

Stecker, P. M., Fuchs, D., & Fuchs, L. S. (2008). Progress monitoring as essential practice within response to intervention. *Rural Special Education Quarterly, 27*(4), 10–17.

Tindal, G., Marston, D., & Deno, S. (1983). *The reliability of direct and repeated measurement* (Research Report No. 109). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Tucker, J. A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52,* 199–204.

University of Oregon Center on Teaching and Learning. (2012). *2012-2013 DIBELS data system update part I: DIBELS next composite score* (Technical Brief No. 1202). Eugene: University of Oregon.

Vannest, K. J., Parker, R., & Dyer, N. (2011). Progress monitoring in grade 5 science for low achievers. *The Journal of Special Education, 44*(4), 221–233.

Wanzek, J., Roberts, G., Linan-Thompson, S., Vaughn, S., Woodruff, A. L., & Murray, C. S. (2010). Differences in relationship of oral reading fluency and high-stakes measures of reading comprehension. *Assessment for Effective Instruction, 35*(2), 67–77.

Wayman, M. M., Wallace, H. I., Wiley, S., Ticha, R., & Espin, C. (2007). Literature synthesis on Curriculum-Based Measurement in reading. *Journal of Special Education, 41,* 85–120.

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* 207–214.

Problem Analysis at Tier 2: Using Data to Find the Category of the Problem

Matthew K. Burns, Kathrin E. Maki, Abbey C. Karich, Matthew Hall, Jennifer J. McComas and Lori Helman

What makes an intervention implemented at tier 2 of a response-to-intervention (RTI) model more intensive than instruction received within tier 1? It could be that a tier 2 intervention occurs in small groups and tier 1 core instruction is geared toward the entire classroom. However, small-group instruction has become a staple of core teaching practices (Allington 2011), and it is not unusual for a small number of students who are receiving similar or the same tier 3 intervention to be grouped together for efficiency. Perhaps it is not the size of the group that matters, but the amount of problem analysis needed to identify the appropriate instructional practice that differentiates a tier 2 intervention from tier 1 instruction and a tier 3 intervention from one delivered for tier 2. Problem analysis is the process by which targeted interventions are identified so that the intervention is directly linked to the nature of the academic need and therefore has a high likelihood of being successful (Tilly 2008). Thus, problem analysis is central to the RTI process and should occur at all three tiers (Christ et al. 2005).

Burns and Gibbons (2012) suggest that the RTI process is essentially answering three problem analysis questions: (a) tier 1—Is there a whole-class problem? (b) tier 2—What is the category of the problem? and (c) tier 3—What is the environmental variable that is most closely related to the problem? These three questions are answered with data and are used to identify appropriate interventions. The questions also represent a continuum of intensity because the question for tier 1 (whole-class problem) is answered with group scores (e.g., median score for a classroom) from universal screening measures, but additional data are needed to answer questions for tiers 2 and 3. The purpose of the current chapter is to describe the problem analysis process for tier 2 with reading and mathematics. First, a review of relevant research is provided, followed by data to support the effectiveness of analyzing problems within tier 2, and concluding with directions for future research.

Research Regarding the Need to Target Tier 2 Interventions

An effective tier 2 is critical to the success of any RTI model. Of course, without quality core instruction, intervention efforts are not likely to be successful, which makes tier 1 the most important component of a school's RTI implementation model. However, an effective tier 2 intervention could support a relatively large number of students and could prevent students from needing even more intensive services (i.e., tier 3). Moreover, a review of research found only low to moderate support for many of the main

M. K. Burns (✉)
University of Missouri, Columbia, MO, USA
e-mail: burnsmk@missouri.edu

K. E. Maki · A. C. Karich · M. Hall · J. J. McComas · L. Helman
University of Minnesota, Minneapolis, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_17

components of an RTI model, but found strong empirical support for providing systematic interventions to target foundational skills in reading (Gersten et al. 2008). Interventions are delivered in small groups to students who score below the benchmark standard on universal screening measures.

Interventions implemented for tier 2 often rely on standardized protocols or commercially prepared intervention packages that tend to be comprehensive in nature (i.e., address multiple components of reading; Vaughn et al. 2008), but that has not always been the case. Some of the earliest RTI implementation efforts relied heavily on individualized interventions that were often derived through a problem-solving team process (e.g., Ikeda and Gustafson 2002; Lau et al. 2006; McNamara and Hollinger 2003). Schools tended to struggle to implement the problem-solving process (Burns et al. 2005b) and doing so consumed too many school resources. For example, consider an elementary school with 650 students. Research has found that approximately 20% of students need support beyond effective core instruction (Burns et al. 2005a), which means that in the hypothetical school with 650 students, approximately 130 of them would need intervention beyond tier 1. If those interventions were developed with a problem-solving team approach, and the problem-solving team met on a weekly basis to discuss two students each time, then there would need to be 65 weeks in the school year to get to all of the students. The 65 weeks would not even include meetings at which the team discusses progress of students who were discussed at previous meetings. There simply is not enough time in the school year to rely on a problem-solving team to develop interventions at tier 2.

Another reason that intervention packages and commercially prepared interventions are used for tier 2 is because there could be some assurance of a research base. There are numerous websites that rate how effective various intervention packages are at improving reading (e.g., http://www.intensiveintervention.org/chart/instructional-intervention-tools). Finally, implementation of a packaged or commercially prepared intervention is likely easier than one that is developed by school personnel, and likely includes materials to assess implementation integrity. However, research might suggest that a comprehensive packaged intervention might not be the most effective approach to deliver tier 2 interventions. Below, the research regarding comprehensive packaged interventions is discussed.

Comprehensive Intervention Packages for Reading

As stated above, small-group interventions can be effective for increasing student skills in reading (Gersten et al. 2008). However, recent research that implemented a standardized small-group intervention addressing word recognition, vocabulary, fluency, and comprehension found only small effects for struggling readers ($d = 0.16$, Vaughn et al. 2010). One potential reason that the effect size for the Vaughn et al. study was smaller than other syntheses of research could be that it was conducted with middle-school students, and most reading intervention research is conducted with elementary-aged students. However, meta-analytic research with adolescent struggling readers found an average effect size that was much larger ($g = 0.95$; Scammacca et al. 2007) than the small effects noted by Vaughn et al. (2010).

Another reason the effect size was small for the Vaughn et al. (2010) study could be that the comprehensive intervention did not adequately target the student needs. Not all types of instruction will be equally beneficial for all students, and in order for students to effectively learn, instruction must be provided that matches their diverse needs (Al Otaiba and Fuchs 2006; Kamps and Greenwood 2005). Juel and Minden-Cupp (2000) found that first-grade students with weaker reading skills benefited more from instruction that was explicit and focused on decoding words, whereas students with higher reading skills benefited more from meaning-based instruction. Similarly, Connor et al. (2009) found that the amount and type of reading instruction necessary for students to achieve proficient reading skills differed for individual students.

Targeted Reading Interventions

A recent meta-analysis compared the effectiveness of a comprehensive intervention (addressed multiple components of reading; $g=0.35$) to a targeted intervention (addressed one component of reading based on student need; $g=0.65$), and found that the latter was more effective than the former (Hall and Burns 2014). Interventions in general were more effective if they targeted the student's area of need (Burns et al. 2008), but how to best accomplish this for small-group interventions has not been well researched. Burns and colleagues (Burns and Gibbons 2012; VanDerHeyden and Burns 2010) proposed a model for tier 2 that uses a standardized approach, but that also targets interventions based on the categories of the National Reading Panel areas (NRP; National Institute of Child Health and Human Development 2000), phonemic awareness, phonics, fluency, and vocabulary/comprehension. Phonological decoding predicted word reading, and the rate and accuracy of word reading predicted comprehension among students who struggled with reading (Berninger et al. 2006). Although the developmental progress of specific reading skills is not linear in nature, assessing how well a student is progressing through them could provide a useful heuristic for most students who experience difficulties. For example, if a student struggled with comprehension and demonstrated adequate reading fluency, then the intervention would focus on comprehension; however, if a student struggled with comprehension, fluency, and decoding, then decoding would be the intervention target because it is the most basic of the three. The model and supporting data are described in more detail below.

Targeted Interventions for Mathematics

Fluent computation is an important goal for mathematics (National Council of Teachers of Mathematics, NCTM 2000; National Math Advisory Panel 2008) and could be a target for small-group interventions because students with difficulties in mathematics often struggle to quickly recall basic mathematics facts (Geary et al. 2007; Hanich et al. 2001). Moreover, students who are not proficient in more advanced mathematics problems often lack fluency in the basic skills within them (Houchins et al. 2004). Students compute fluently when they solve mathematics problems more quickly if they recall the answer rather than perform the necessary mental algorithm (Logan et al. 1996). For example, fluent computation can occur when a student can look at $5\times6=$ and quickly recall that the answer is 30 without counting by 5's or some other manual computation. Providing additional practice with basic or component skills (e.g., single-digit multiplication) has consistently led to increased performance of the more advanced skills (Dehaene and Akhavein 1995; Singer-Dudek and Greer 2005). Thus, tier 2 interventions for mathematics tend to focus on building fluency of the basic skill.

Problem analysis for tier 2 in mathematics focuses on identifying the correct skill to target fluency building, which is best accomplished by a series of single-skill survey assessments. The analysis begins by sequencing the skills or objectives within a mathematics curriculum (e.g., single-digit multiplication, then single-digit division, then multi-digit multiplication, etc.). Most mathematics curricula provide this sequence of objectives. Next, a series of single-skill curriculum-based assessments are created to represent each skill or objective. There are many free websites available with which single-skill probes can be created including www.mathfactcafe.com, www.aplusmath.com, and www.interventioncentral.com. After creating the probes, the sequence of the skills within the curriculum will determine the sequence with which the probes are used. Students are assessed for 2 or 4 min each depending on the skill being assessed.

After the data are collected, they are converted to digits correct per minute and are compared to instructional-level criteria to find the highest skill from the survey assessments in which the student scores within an instructional level, and intervention begins with that skill. Deno and Mirkin (1977) provide instructional-level criteria for mathematics that are commonly used, but those criteria were derived from experience in one school in Minnesota (S. L. Deno, personal communication, April 15, 2005) and were not based on research. Burns et al. (2006) empirically derived instructional-level criteria and found that 14–31 digits correct per minute (dcpm) for second and third graders and 24–49 dcpm for fourth- and fifth-grade students represented an appropriate level of challenge for mathematics. Thus, single-skill assessments are administered, usually in reverse order, until the student scores within the instructional-level range.

Research has consistently demonstrated the effectiveness of using survey-level assessment to identify the mathematics objective that represented an instructional level for an individual student and intervening with that skill (Burns et al. 2010; Spicuzza et al. 2001; Ysseldyke et al. 2003). Moreover, previous research used flashcards with student dyads that were matched with instructional-level data in order to build fluency of basic mathematics skills, which resulted in significant gains in mathematics computation skills (VanDerHeyden and Burns 2005).

Effectiveness of Targeting Tier 2 Interventions for Reading

The current data were collected during the second year of a 3-year partnership (Path to Reading Excellence in School Sites; PRESS) among six urban schools, a research university, a statewide service organization, and a national corporation. PRESS is a comprehensive research-based approach to early literacy that was designed to prepare all students to read at grade level by the end of third grade.

Data presented below were taken from four of the six participating PRESS schools that were traditional public schools from one urban district in Minnesota. There were a total of 316 second-grade students and 315 third-grade students across the four elementary schools. The total sample consisted of 51.4% females and 14% white students (86% from a minority background), and 80% were eligible for the federal free or reduced lunch program. There were 10 students in second grade who received special education services for reading and 12 in third grade. Thus, the total number of students represented in the data below was 306 for second grade and 303 for third.

Measures

Benchmark: Oral Reading Fluency All second- and third-grade students were assessed with oral reading fluency (ORF) measures from AIMSweb (Pearson 2008) in the fall, winter, and spring of the academic school year as a universal screener. Data were recorded as the number of words read correctly (WRC) during each 1-min assessment. The ORF scores ranged from 0 to 181 (M = 59.68, SD = 47.80) in the fall for second grade, 2 to 194 (M = 97.72, SD = 44.64) for spring of second grade, 0 to 232 (M = 76.63, SD = 44.56) for fall of third grade, and 7 to 271 (M = 105.91, SD = 47.88) for spring of third grade. Students who scored below the seasonal benchmark associated with the system (Pearson 2008) were identified as struggling readers and received a tier 2 intervention. In addition, ORF data were used to assess progress by computing a slope across the three benchmark assessments using weeks within an ordinary least-square calculation.

Benchmark Measures of Academic Progress In addition to the ORF seasonal benchmark screener, students were assessed with the Measures of Academic Progress (MAP) assessment (Northwest Evaluation Association 2003) for reading three times per year. MAP Reading is a norm-referenced computer adaptive test that is designed to measure growth across a year or several years. The second- and third-grade MAP reading measure assesses several areas of student comprehension, including word analysis, vocabulary, literal comprehension, interpretive comprehension, and literary response and analysis. The MAP assessment was therefore used as a global measure of reading comprehension.

Students completed the MAP on a computer in one 40–60-min session. MAP assessment scores are presented in Rasch units (i.e., RIT scores), allowing for comparisons across grades. Students scoring at or below the 25th percentile struggled with comprehension. The scores ranged from 136 to 213 (M = 169.28, SD = 17.82) in the fall for second grade, 142 to 223 (M = 181.05, SD = 16.90) for spring of second grade, 144 to 234 (M = 183.51, SD = 18.02) for fall of third grade, and 139 to 234 (M = 190.49, SD = 19.12) for spring of third grade. MAP benchmark scores were converted to a rate of growth using weeks within ordinary least squares.

Progress Monitoring: ORF Although not presented here, student progress was monitored on a regular basis using ORF from the Formative Assessment System for Teachers (FAST; Christ et al. 2011). All students were assessed every other week with grade-level ORF measures. Data were converted to a slope using ordinary least squares to represent average growth per week. The slope estimates were compared to criteria for slopes based on rate of growth needed to obtain seasonal benchmark standards. The fall, winter, and spring benchmark criteria for both measures were used to compute slope estimates with ordinary least squares. For example, AIMSweb (Pearson 2008) criteria indicate that a student in second grade should read 50 WRC per minute in the fall, 80 in the winter, and 92 in the spring. Assuming that those data are collected in the 2nd, 18th, and 34th weeks of the year, the resulting slope would be 1.28 WRC per minute per week. Students whose slope was at least 1.28 were making sufficient progress during the intervention. The third-grade ORF benchmark criteria were 90, 91, and 109, which resulted in a slope of 1.22 WRC per minute increase per week. Third-grade students who demonstrated a slope of at least 1.22 were considered to be making sufficient progress during intervention.

Diagnostic Assessment Process

The model articulated by Burns and colleagues (Burns and Gibbons 2012; VanDerHeyden and Burns 2010) in which tier 2 consists of standardized reading intervention targeted toward phonemic awareness, phonics, fluency, or vocabulary/comprehension was used to target the tier 2 intervention. The diagnostic model is presented in Fig. 1.

All students were screened for reading interventions with MAP and ORF. Students who scored below the 25th percentile were identified as struggling with comprehension, which was the first step in the diagnostic model. If the student demonstrated low comprehension, but adequate reading fluency (an ORF score at or above seasonal benchmark), then comprehension was the most fundamental skill in which the student struggled and the student received a comprehension intervention. If the student's ORF score was below the seasonal benchmark, then the student demonstrated a fluency deficit. However, in the latter scenario, the more fundamental skill of decoding was assessed and ruled out as an intervention target.

Decoding was screened by examining the accuracy with which students read the words during the ORF assessments. Reading accuracy is computed as the percentage of words read correctly (number of words read correctly/number of total words), which results in reliable data (Burns et al. 2000) that can be useful for instructional decision-making (Burns 2007; Hosp and Ardoin 2008; Treptow et al. 2007). The percentage of words read correctly is compared to the research-based criterion of 93 % or higher (Gickling and Armstrong 1978; Treptow et al. 2007). Thus, students who read less than 93 % of the words correctly are likely struggling to decode the text and would likely benefit from reading instruction (Burns and Parker in press). There may be other measures of decoding that work well too, but examining accuracy data is likely sufficient to hypothesize the intervention target for most students. It may be beneficial to further screen decoding skills of students who score close to 93 % correct with any one of a number of assessments including a nonsense word fluency measure, or a word-attack subtest from a standardized norm-referenced reading assessment.

Phonemic awareness was not routinely screened for students in second and third grade, *Diagnostic Assessment Process Used to Target Interventions*

Step	Data	Criterion	Conclusion
Step 1	Examine MAP data.	Is score at or below the 25th percentile?	Yes = Examine fluency (Step 2). No = Student does not need intervention.
Step 2	Examine ORF data.	Is score at or below the benchmark criterion?	Yes = Examine decoding (Step 3). No = Provide an intervention that targets comprehension.
Step 3	Examine accuracy data.	Did student read less than 93% of the words correctly?	Yes = Examine phonemic awareness (Step 4). No = Provide an intervention that targets decoding.
Step 4	Examine phonemic awareness data.	Does the student struggle with phonemic awareness?	Yes = Provide an intervention that targets phonemic awareness. No = Provide an intervention that targets decoding/phonics.

Note – MAP = Measures of Academic Progress for Reading, and ORF = Oral Reading Fluency.

Fig. 1 Diagnostic assessment process used to target interventions

but was screened with students in kindergarten and first grades with phoneme segmentation fluency from AIMSweb (Pearson 2008). Phonemic awareness is the knowledge that words are made of individual sounds and that those sounds can be manipulated to make new words, and it is usually well developed by second grade (National Institute of Child Health and Human Development 2000). However, the phonemic awareness of second- and third-grade students who were suspected to have a difficulty with this most fundamental of all reading skills was screened using the Quick Phonemic Awareness Assessment (QPAA; PRESS Research Team 2013a). The authors did not have a consistent decision rule to determine if phonemic awareness should be assessed and relied on either teacher-generated data or reading accuracy scores that were quite low (e.g., less than 80 %). The QPAA is a 20-item assessment that examines rhyming, blending, segmenting, and initial sound with five items in each area. Students are asked to generate an answer to each item (e.g., "tell me a word that rhymes with [given word]" or "I am going to say some sounds, what word to you hear when you say those sounds fast?"). The test stops if the student does not correctly respond to any two items within one area, and failure to complete the test suggests a phonemic awareness deficit. A student with low decoding skills but sufficient phonemic awareness received a decoding intervention, and those with low phonemic awareness participated in phonemic awareness intervention.

Grade-Level Teams Student screening and diagnostic data were interpreted by the schools' grade-level teams (GLTs), which were teams of teachers made up of all of those who taught a particular grade level. The GLT process was based on the professional learning community model (DuFour et al. 2005; Hord 1997) because of the focus on student outcome data and creating a culture of collaboration to enhance student learning (DuFour 2005). The GLTs met on a weekly basis for 60 or 90 min.

One of the monthly GLT meetings in September, January, and May focused on examining universal screening data and relied on an agenda and analysis-to-action form that embedded the following questions: (a) Is there a whole-class problem? (b) Who needs a tier 2 intervention? (c) Among students needing a tier 2 intervention, what is the category of the problem for each? and (d) Are there any students for whom we should go immediately to a tier 3 intervention? It should be noted that there were very few students receiving a tier 3 intervention across the PRESS project, with less than 10 in second and third grade from each of these four schools.

The GLT meetings that examined screening data were facilitated by a data manager who was well trained in the screening and diagnostic process. Each school was supported by a literacy coach who was trained in the PRESS model and by one school psychology graduate student with advanced training in diagnostic and general assessments. The coach or graduate student facilitated the meeting by projecting student data and having the team answer the four questions outlined above using an analysis-to-action form. Each teacher completed an analysis form for his or her students, but the process was completed as a group.

Interventions

There were three conditions in which students participated. A total of 175 second- and third-grade students scored below the seasonal benchmark criterion and received an intervention in a small group from the PRESS program that targeted their reading deficit. An additional 69 students also fell below the benchmark criteria, but received a comprehensive intervention that was delivered by school personnel. Finally, there were 365 students in second and third grade who scored at or above seasonal benchmark criteria and received only core instruction with no additional interventions.

Targeted Intervention Students in the first group were identified as needing a tier 2 intervention and received a PRESS intervention (PRESS Research Team 2013b) that targeted the most basic skill in which the student struggled. Interventions were delivered in a small group of two to five students for approximately 20 min per day for 4 days per week. The interventionists were graduate students in school psychology, special education, or curriculum and instruction. The interventions are succinctly described below, but further information and implementation protocols/checklists are available at http://www.cehd.umn.edu/reading/PRESS/resources/interventions.html.

Targeted Phonemic Awareness Interventions Phonemic awareness interventions focused on the isolation and identification of sounds as well as sound manipulation in words. Eight isolation/identification interventions were used with students unable to identify individual phonemes in words. Isolation/identification interventions 1 and 2 required students to identify initial sounds in words, interventions 3 and 4 required students to identify final sounds in words, the interventions 5 and 6 focused on the identification of the middle sounds in words, and the final two isolation/identification interventions required students to sort picture cards based on initial, middle, and final sounds.

In addition to the isolation/identification phonemic awareness interventions, five manipulation interventions were implemented with students able to isolate sounds in words, but who could not manipulate sounds within words. In manipulation intervention 1, students deleted initial and final sounds from words. In manipulation intervention 2, students substituted initial sounds in words. Phonemic awareness manipulation intervention 3 required students to substitute final sounds, and manipulation intervention 4 required students to substitute medial sounds. The fifth manipulation intervention required students to substitute initial, final, and medial sounds.

Targeted Phonics Interventions Six phonics interventions were implemented with students who had acquired phonemic awareness, but who lacked sufficient decoding skills. In interventions 1 and 2, students were explicitly taught letter–sound correspondence by matching picture cards with letters. Intervention 3 focused on letter–sound correspondence through the use of Elkonin boxes (1971). Students placed magnetic letters in the boxes based on where the sound was heard in a word. The fourth intervention elicited word-building activities by manipulating graphemes within words. Students again used Elkonin boxes in this intervention by manipulating one grapheme within a word to create a new word. In the fifth phonics intervention, more advanced phonics skills such as vowel patterns were the focus, through word writing and passage reading. Students wrote words on white boards, focusing on a specific phonics skill; then, students identified words with the targeted grapheme or letter combination while reading a passage. In the final phonics intervention, students analyzed words by sorting word cards into three categories based on graphemes or combinations of letters.

Targeted Fluency Interventions Two fluency interventions were implemented with students who struggled with speed, accuracy, and/or expression in reading, but had mastered phonemic awareness and phonics skills. In the supported cloze procedure (Rasinski 2003), students read an instructional-level passage, working in pairs with a peer or with a graduate research assistant, alternating reading every other word while providing error correction. Pairs read the passage three times total, for 1 min each time. An interventionist also provided assistance or error correction as needed.

Students needing additional support in rate and expression participated in the repeated reading intervention (Samuels 1979). Students read a passage three times, for 1 min each time. The interventionist provided error correction as needed at the conclusion of each reading. After the second and third readings, the interventionist asked the student comprehension questions including what the passage was mostly about and the most important information in the passage. After the three independent readings, the student and interventionist finished reading the story together.

Targeted Comprehension The comprehension intervention was based on reciprocal teaching (Palinscar and Brown 1984), which was used to teach the following comprehension strategies: prediction, summarization, question generation, and clarifying, along with activating prior knowledge. The interventionists used grade-level texts to teach these skills and had the students complete comprehension questions about them to determine if the skills were being learned.

Comprehensive Intervention Each of the schools also chose to implement a different and more comprehensive intervention for a portion of students through their school-based supplemental service (e.g., Title 1). The schools used the Fountas and Pinnell Leveled Literacy Intervention System (LLI; Fountas and Pinnell 2011). The LLI is a small-group, supplementary literacy intervention that focuses on comprehension and fluency, with decoding and phonemic awareness to be embedded throughout the lessons. Examples of strategies taught with LLI include reading aloud, writing, phonics/word study, attention to features of genre, attention to disciplinary reading, literature inquiry, writing about reading, close reading, and more. LLI was delivered by school personnel three to five times each week, with three to five students in each group. There was no consistently implemented decision rule to determine how the schools selected the students to receive the LLI. For example, students were selected in one school because they were judged to have the strongest skills among those identified as needing support, but in another school they were selected based on eligibility for Title 1 and teacher nomination. A total of 38 second-grade students and 31 third-grade students received the comprehensive intervention.

Tier 1 Students who scored above screening benchmark criteria received no supplemental intervention. They participated only in the core classroom literacy instruction and served as a control group for this demonstration. A total of 188 second-grade and 177 third-grade students fell within this category.

Core Instruction

The four schools used a balanced literacy approach to reading instruction that included guided reading groups based on the Fountas and Pinnell (1996) reading program. All participating classrooms were observed multiple times throughout the school year with a 54-item observation protocol to assess the quality of the core instruction. The scale included 28 items that addressed the classroom environment (e.g., accessibility of writing tools and varied reading materials) and 26 items that addressed the instructional practice (e.g., explains purpose of the lesson, fosters discussion). Each item was rated on a 0–3 scale with a 3 indicating that the item was implemented at an expert level. The mean rating for the classrooms on the final observation, which occurred in May, was 1.82 for environment items and 2.01 for instructional practice. Thus, overall, the schools seemed to use effective instructional practice within the reading curriculum used for core instruction.

Intervention Integrity

The interventionists were trained in the implementation of the interventions and relevant assessments during one 3-h training at the beginning of the year. Each followed a protocol that included scripted administration procedures for each intervention. Before beginning interventions, all interventionists were assessed on their ability to implement each intervention accurately, and intervention sessions were observed on multiple occasions throughout the year to ensure fidelity. All observed interventions resulted in at least 90 % fidelity to the model. However, no fidelity data were available for the comprehensive (LLI) intervention.

Education graduate students working for the PRESS project collected all data. The ORF administration procedures were evaluated before the graduate students began collecting data. A second assessor recorded WRC, and all words that were consistently rated as correct or incorrect across both observers were counted as agreements. Inconsistent ratings were counted as disagreements. The total number of agreements was divided by the total number of words and multiplied by 100 to obtain interobserver agreement (IOA). All data collectors demonstrated at least 95 % IOA before they began collecting data.

Analyses

Growth was evaluated using a multivariate analysis of covariance (MANCOVA) in which ORF and MAP slope from the three benchmark assessments served as the dependent variables. Student ORF and MAP growth served as the outcome variables because both measure student reading skills, but ORF measures only fluency whereas MAP measures comprehension. The fall ORF benchmark score served as the covariate.

Growth was also compared to a criterion based on 1 year's worth of growth. Second-grade students with a slope of at least 1.28 WRC per minute per week were identified as having made at least 1 year's worth of growth, and third graders at or above 1.22 WRC per minute per week were identified as having made 1 year's worth of growth.

Results

The project examined the differential effectiveness of targeted and comprehensive tier 2 reading interventions. Descriptive statistics regarding the mean ORF and MAP growth for students participating in targeted and comprehensive interventions as well as students not receiving tier 2 interventions are presented in Table 1. Second-grade students participating in targeted tier 2 interventions mean growth per week was 1.33 words read correctly per minute (WRCM) whereas second-grade students not receiving supplemental tier 2 intervention had a mean weekly growth of 1.25 WRCM. Moreover, students participating in comprehensive tier 2 interventions (i.e., not directly targeted toward student skill need) had an average weekly growth of 1.07 WRCM. Third-grade students participating in targeted tier 2 interventions mean weekly growth was 1.23 WRCM compared to 1.03 WRCM for students not participating in tier 2 interventions and 0.94 WRCM for students receiving a comprehensive tier 2 intervention.

A MANCOVA comparing student ORF growth and MAP growth was conducted to compare the progress of students receiving targeted and comprehensive tier 2 interventions. Findings showed that second- and third-grade students receiving targeted tier 2 interventions made statistically significant greater growth than students receiving comprehensive interventions or no tier 2 interventions. Partial η^2 was used as an estimate of effect size and resulted in a moderate to large effect for second grade ($\eta^2 = 0.12$) and a large effect for third grade ($\eta^2 = 0.16$).

The percentages of students making at least 1 year's reading growth on the ORF and MAP measures are presented in Table 2. A total of 70 % of second-grade students in targeted interventions made at least 1 year of growth compared to 62 % of students receiving no tier 2 intervention and 55 % of students receiving a comprehensive intervention. In third grade, 73 % of students participating in the targeted tier 2 interventions made at least 1 year's worth of growth compared to 70 % of students receiving no tier 2 intervention and 48 % of students receiving a comprehensive intervention.

Implications for Practice and Research

Students participating in targeted tier 2 interventions made significantly more growth over one school year than students participating in comprehensive tier 2 interventions and students receiving no tier 2 intervention. Thus, targeting interventions directly to students' area of need

Table 1 Mean ORF and MAP growth in average increase per week for second- and third-grade students in targeted intervention, comprehensive intervention, and tier 1

	Targeted	Comprehensive	Tier 1
Second grade	$N = 80$	$N = 38$	$N = 188$
ORF	1.33	1.07	1.25
MAP	0.56	0.40	0.39
Third grade	$N = 95$	$N = 31$	$N = 177$
ORF	1.23	0.94	1.03
MAP	0.37	0.42	0.35

MAP measures of academic progress for reading, *ORF* oral reading fluency
Targeted intervention grade 2 $\Lambda\,(3, 302) = 0.71$, $p < 0.05$
Targeted intervention grade 3 $\Lambda\,(3, 299) = 0.77$, $p < 0.05$

Table 2 Percentage of students making 1 year's growth on CBM-R and/or MAP

	Tier 1		Targeted		Comprehensive	
	Neither measure (%)	At least 1 measure (%)	Neither measure (%)	At least 1 measure (%)	Neither measure (%)	At least 1 measure (%)
Second grade	38	62	30	70	45	55
Third grade	30	70	27	73	52	48

CBM-R Curriculum-Based Measurement Reading, *MAP* Measures of Academic Progress for reading, *ORF* oral reading fluency resulted in greater growth for both second- and third-grade students, which was consistent with previous research emphasizing the need to target intervention efforts (Burns et al. 2008).

The average rates of growth for the targeted interventions were higher than the criteria based on AIMSweb (Pearson 2008) benchmark standards, but the rate of growth for students not receiving intervention did not represent 1 year's worth of growth. Thus, the students receiving targeted interventions narrowed the gap with their peers and potentially with grade-level benchmark standards as well. The lower average rate of growth for the comprehensive intervention suggested that those students could have fallen further behind their peers and standards. Further implications for practice are presented below and are summarized in Table 3.

Table 3 Summary of implications for practice

Tier 2 interventions are an important part of a comprehensive RTI framework
Providing targeted reading interventions to struggling readers was more effective than using a comprehensive model
Phonemic awareness, phonics, fluency, and vocabulary/comprehension served as categories of reading problems to group students for intervention purposes
Intervention addresses the most fundamental skill with which the student struggles
Identifying categories of reading problems provides a low-level problem analysis that balances individualizing interventions with the advantages of a standardized approach
Grade-level teams interpret screening data three times each year to answer: (a) Is there a whole class problem? (b) Who needs a tier 2 intervention? (c) Which intervention is most appropriate based on the category of the problem? And (d) is there any student for whom a tier 3 intervention is immediately warranted?
Teachers must be well trained in the use of assessment data to design and monitor interventions
Schools can enhance the ability of grade-level teams to interpret data with an easy-to-use data warehouse system and a data manager to facilitate the meeting
Implementation integrity of the interventions is important to assess and monitor
Without good core instruction, nothing else matters

RTI response to intervention

Implications for Practice

Tier 2 interventions are an important part of a comprehensive RTI framework (Marston 2005; Reschly 2008). RTI models are preventative, data-driven service delivery models, which provide intervention to students struggling with academic skills. Moreover, RTI models utilize assessment to identify struggling readers and to provide explicit instruction in the student's area of need (Fuchs and Fuchs 2006; Justice 2006). Thus, providing targeted reading interventions to struggling readers is an effective method for increasing students' reading skills.

The term "targeted" might have different meanings depending on who is asked. Many schools refer to their tier 2 interventions as "targeted" because they are provided at the targeted level in the universal–targeted–indicated continuum (or tier 3 in the universal–selected–targeted continuum; Weissberg and Greenberg 1998) or because the students identified as below the benchmark have been "targeted" for intervention. However, the authors suggest that "targeted" should refer to more than just how many students receive the intervention and should describe the focus of intervention efforts. The NRP (National Institute of Child Health and Human Development 2000) categories were used to precisely target the area of need for students participating in targeted tier 2 interventions whereas the comprehensive intervention group all received the same intervention regardless of their area of need. Previous research has shown that, for struggling readers, code-based skills such as phonics and fluency are prerequisites to reading comprehension (Berninger et al. 2006). Thus, targeting intervention toward those code-based skills for struggling readers is necessary to ensure that students have the necessary foundational skills.

The model suggested by Burns and colleagues (Burns and Gibbons 2012; VanDerHeyden and Burns 2010) and used here within the PRESS model seems to provide an effective balance between individualizing and standardizing interventions. As stated above, many schools use standardized commercially prepared interventions because they are easy to obtain, are easy to implement, and have a research base. All of the interventions used in this study were implemented with high fidelity, and utilized components or intervention approaches with a considerable research base. However, each was identified with a low-level analysis that mostly involved examining existing data. Most school personnel would agree that it would be ideal to individualize interventions for every student who needs support, but cannot do so for logistical reasons. The current model allowed for a broad categorical approach to individualize interventions, which seemed to be more effective than using the same comprehensive intervention for every student.

These data suggest an effective intervention system, but the positive outcomes are dependent on several factors such as high implementation integrity and adequate school attendance. Perhaps the most important aspect of the model was the presence of a data manager or coach to facilitate the meeting. The current data were taken from the 2nd year of the 3-year PRESS project. All of the teachers in the participating schools were instructed in screening, diagnostic decisions, and monitoring student progress during the 1st year. New staff was also trained during the 2nd year and teachers were provided booster sessions regarding the uses of the assessment data. Moreover, the teachers were shown how to implement the interventions during the 2nd year because implementation was turned over to the teachers during the 3rd year of the project. Thus, the teachers were trained in using data to identify students who need additional support and to determine what they needed. However, it is likely that they would have struggled to implement those skills three times each year without the support of a guide who was well trained in the process and without an effective data management system.

It is considered wise to dedicate professional learning time to instruct staff on the use of data, and provide extended time from one person to facilitate these data meetings. The school psychologist could be ideally suited to facilitate the three meetings each year in which screening data are discussed, but it would be difficult for one person to attend all six meetings (assuming there is one GLT for every grade at an elementary school) after each seasonal benchmark. Thus, it might be beneficial to have two people who have expert-level understanding to facilitate these meetings.

Another important consideration is the quality of the core instruction. The PRESS project involved assessing the core instruction and using the data to provide coaching around literacy instruction. Thus, there was some assurance that aspects of quality core instruction occurred. However, there was certainly variability in the quality of the core instruction, which was problematic. The model that was used here targets reading interventions and focuses intervention efforts on that particular goal. This approach is only effective if students receive effective core instruction. In other words, reading can be broken apart and remediated, only if it is being reassembled somewhere else with quality balanced instruction. It seems that part of the allure of comprehensive interventions could be that they have the potential to fill holes in core instruction. Our suggestion would be to target the interventions and to provide effective balanced instruction rather than filling holes elsewhere.

Implications for Research

Previous research has shown that targeted reading interventions result in moderate reading growth ($g=0.52$; Piasta and Wagner 2010) that was larger than comprehensive small-group interventions ($g=0.65$ and 0.35, respectively; Hall and Burns 2014). However, the comparison was conducted through meta-analytic research and the current project was not experimental in nature. Thus, additional research is needed to determine if it is more effective to target an intervention or to use a commercially prepared comprehensive reading intervention.

Although the findings from the PRESS project provide evidence that directly targeting a student's area of need results in more growth than providing comprehensive interventions, the results should be interpreted within their limitations. First, students were not randomly assigned to targeted, comprehensive, or no intervention groups. This was not designed to be an experimental study, but was a demonstration and an implementation project with supporting data. Thus, threats to internal validity cannot be excluded, such as differences between participants across groups. Moreover, extraneous interfering variables impacting the results cannot be ruled out. Second, intervention fidelity of implementation was not assessed for students participating in the comprehensive tier 2 interventions (LLI). Therefore, smaller growth for students participating in comprehensive interventions could be affected by a potential lack of implementation fidelity.

Future Directions

Because the current study did not utilize random assignment of groups, future research should replicate the current study with random group assignment. Such research would provide stronger evidence regarding the differential effectiveness of targeted and comprehensive tier 2 interventions. The results of the current study suggest that targeting tier 2 interventions to students' area of challenge is effective, but how best to implement such targeted interventions is not well established within the literature. Therefore, future research should examine how best to intervene in small groups.

There is also considerable research needed to determine the most effective approach to conducting a GLT meeting. Many schools are using the professional learning community (Dufour et al. 2005) approach to teaming within and across grade levels, but the research supporting that approach is limited. Moreover, it is unknown how well teams actually implement the model that they reportedly use.

The diagnostic assessment model used by PRESS was supported with earlier research, but has not been tested empirically beyond examining student outcomes. Greater student outcomes in comparison to a control group of students not needing intervention and to one that consisted of students receiving a different intervention make an argument for the validity of the model (Kane 2013; Messick 1995), but additional research is needed. For example, the diagnostic accuracy of the assessment system could be evaluated by comparing the results to other measures of the same construct (e.g., Does reading less than 93 % of the words correctly identify the same students as needing decoding help as a different measure of reading decoding?).

Conclusion

The current chapter found that targeting tier 2 interventions was effective and provided a model to do so. Problem analysis is, by definition, the process of finding interventions with a high likelihood for success (Tilly 2008), but it seems that most school personnel do not engage in this process before beginning tier 2 interventions. There is considerable future research to be conducted, but the process outlined here was brief, did not involve many additional resources, could be completely embedded within an existing RTI framework, and improved student learning. The model used by the PRESS project is just one approach, and there may be many others. However, no others have likely been evaluated in comparison to an active intervention with such a large group of students. Thus, additional research is needed, but the research seems warranted.

References

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39,* 414–431.

Allington, R. (2011). What at-risk readers need. *Educational Leadership, 68*(6), 40–45.

Berninger, V. W., Abbott, R. D., Vermeulen, K., & Fulton, C. M. (2006). Paths to reading comprehension in at-risk second-grade readers. *Journal of Learning Disabilities, 39,* 334–351.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School Psychology Quarterly, 22,* 297–313.

Burns, M. K., & Gibbons, K. (2012). *Response to intervention implementation in elementary and secondary schools: Procedures to assure scientific-based practices* (2nd ed.). New York: Routledge.

Burns, M. K., & Parker, D. C. (in press). Using the instructional level as a criterion to determine student phase of learning for reading fluency: Evidence for the learning hierarchy. *Reading and Writing Quarterly*.

Burns, M. K., Tucker, J. A., Frame, J., Foley, S., & Hauser, A. (2000). Interscorer, alternate–form, internal consistency, and test–retest reliability of Gickling's model of Curriculum–based Assessment for reading. *Journal of Psychoeducational Assessment, 18,* 353–360.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005a). Meta-analytic review of responsiveness- to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Burns, M. K., Vanderwood, M., & Ruby, S. (2005b). Evaluating the readiness of prereferral intervention teams for use in a problem-solving model: Review of three levels of research. *School Psychology Quarterly, 20,* 89–105.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35,* 401–418.

Burns, M. K., VanDerHeyden, A. M., & Boice, C. H. (2008). Best practices in intensive academic interventions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1151–1162). Bethesda: National Association of School Psychologists.

Burns, M. K., Klingbeil, D. A., & Ysseldyke, J. E. (2010). The effects of technology-enhancedformative evaluation on student performance on state accountability math tests. *Psychology in the Schools, 47,* 582–591.

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to- intervention vernacular: Clarifying meaningful differences. *Communiqué, 34*(3), 1–2.

Christ, T. J., Ardoin, S. A., & Eckert, T. (2011). *Formative Assessment and Instrumentation for Reading (FAIP-R): Curriculum based measurement for reading*. Minneapolis: University of Minnesota.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., & Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development, 80,* 77–100.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 314–326.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston: Council for Exception Children.

DuFour, R. (2005). What is a professional learning community. In R. DuFour, R. Eaker, & R. DuFour (Eds.), *On common ground: The power of professional learning communities* (pp. 31–44). Bloomington: Solution Tree.

DuFour, R., Eaker, R., & DuFour, R. (2005). Recurring themes of professional learning communities and the assumptions they challenge. In R. DuFour, R. Eaker, & R. DuFour (Eds.) *On common ground: The power of professional learning communities* (pp. 7–30). Bloomington: Solution Tree.

Elkonin, D. (1971). Development of speech. In A. V. Zaporozhets & D. B. Elkonin (Eds.), *The psychology of preschool children*. Cambridge: M.I.T. Press.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth: Heinemann.

Fountas, I. C., & Pinnell, G. S. (2011). *Leveled literacy intervention*. Portsmouth: Heinemann.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93–99.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1359.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009–4045)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gickling, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprehension. *Journal of Learning Disabilities, 11,* 559–566.

Hall, M., & Burns, M. K. (2014). *A meta-analysis of small-group reading interventions. Paper submitted for publication*. Minneapolis: University of Minnesota.

Hanich, L.B., Jordan, N.C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93,* 615–626.

Hord, S. (1997).) *Professional learning communities: Communities of continuous inquiry and improvement*. Austin: Southwest Educational Development Laboratory. http://www.sedl.org/pubs/change34/.

Hosp, J. L., & Ardoin, S. P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention, 33,* 69–77.

Houchins, D. E., Shippen, M. E., & Flores, M. M. (2004). Math assessment and instruction for students at-risk. In R. Colarusso & C. O'Rourke (Eds.), *Special education for all teachers* (3rd ed.). Dubuque: Kendall/Hunt Publishing.

Ikeda, M. J., & Gustafson, J. K. (2002). *Heartland AEA 11's problem solving process: Impact on issues related to special education*. Research report no. 2002-01. Johnston: Heartland Area Education Agency 11.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35,* 458–492.

Justice, L. M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools, 37,* 284.

Kamps, D., & Greenwood, C. R. (2005). Formulating secondary-level reading interventions. *Journal of Learning Disabilities, 38,* 500–509.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.

Lau, M. Y., Sieler, J. D., Muyskens, P., Canter, A., Vankeuren, B., & Marston, D. (2006). Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large midwest urban school district. *Psychology in the Schools, 43,* 117–127.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 620–638.

Marston, D. (2005). Tiers of intervention in responsiveness to intervention: Prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities, 38,* 539–544.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

McNamara, K., & Hollinger, C. (2003). Intervention-based assessment: Evaluation rates and eligibility findings. *Exceptional Children, 69,* 181–194.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: Author.

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the national math advisory panel*. Washington, DC: U.S. Department of Education.

Northwest Evaluation Association. (2003). *Measures of academic progress*. Lake Oswego: Author.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175.

Pearson. (2008). *Aimsweb: Progress monitoring and RtI system*. Bloomington: author.

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly, 45,* 8–38.

PRESS Research Team. (2013a). *PRESS assessment manual*. Minneapolis: University of Minnesota, Minnesota Center for Reading Research.

PRESS Research Team. (2013b). *PRESS intervention manual*. Minneapolis: University of Minnesota, Minnesota Center for Reading Research.

Rasinski, T. V. (2003). *The fluent reader*. New York: Scholastic.

Reschly, D. J. (2008). School psychology paradigm shift and beyond. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V*, *1*, (pp. 3–15), Bethesda: National.

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 50*(5), 376–381.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., et al. (2007). *Intervention for adolescent struggling readers: A meta-analysis with implication for practice*. Portsmouth: RMC Research Corporation, Center on Instruction.

Singer-Dudek, J., & Greer, R. D. (2005). Long-term analysis of the relationship between fluency and the training and maintenance of complex math skills. *The Psychological Record, 55*, 361–376.

Spicuzza, R., Ysseldyke, J. E., Lemkuil, A., McGill, S., Boys, C., & Teelucksingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39,* 521–542.

Tilly, W. D. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 17–36). Bethesda: National Association of School Psychologists.

Treptow, M. A., Burns, M. K., & McComas, J. J. (2007). Reading at the frustration, instructional, and independent levels: Effects on student time on task and comprehension. *School Psychology Review, 36,* 159–166.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30*(3), 15–29.

VanDerHeyden, A. M., & Burns, M. K. (2010). *Essentials of response to intervention*. New York: Wiley.

Vaughn, S., Fletcher, J. M., Francis, D. J., Denton, C. A., Wanzek, J., Wexler, J., & Romain, M. A. (2008). Response to intervention with older students with reading difficulties. *Learning and Individual Differences, 18,* 338–345.

Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. D., Barth, A., Romain, M., & Francis, D. J. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 39,* 3–21.

Weissberg, R., & Greenberg, M. T. (1998). Community and school prevention. In I. Siegel & A. Renninger (Eds.), *Handbook of child psychology, Child psychology in practice* (5th ed., pp. 877–954). New York: Wiley.

Ysseldyke, J. E., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk, 8*(2), 247–265.

Multilevel Response-to-Intervention Prevention Systems: Mathematics Intervention at Tier 2

Lynn S. Fuchs, Douglas Fuchs and Amelia S. Malone

As with reading, poor mathematics learning in school is associated with serious, lifelong difficulties (e.g., National Mathematics Advisory Panel (NMAP) 2008; Rivera-Batiz 1992), and the prevalence of mathematics difficulty is as high, with 5–9% of the population experiencing such problems (e.g., Dirks et al. 2008; Shalev et al. 2000). Moreover, although the prevalence and lifelong debilitating consequences are similar for learning disabilities in mathematics and reading, mathematics has received much less emphasis than reading. This is the case not only in research but also in school practice. That is, schools are much more likely to provide Tier 2 intervention in reading than in mathematics.

At the same time, mathematics difficulty may require more attention than reading difficulty in both research and schools. This is because mathematics difficulty is potentially more complicated and difficult to address, given that the mathematics curriculum is organized into many more strands that are presumed to represent different component skills. In reading, measurement studies (e.g., Mehta et al. 2005) provide the basis for five-component reading skills: phonological awareness, decoding, fluency, vocabulary, and comprehension. In mathematics, measurement studies are yet to be conducted, but the assumption reflected in the curriculum is that many more component skills exist. For example, just considering the elementary school grades, one major focus is whole numbers, which is subdivided into curricular strands: concepts, numeration, basic facts, algorithmic computation, and word problems. Another major focus is fractions, which includes common fractions, decimals, and proportions and has its own set of subdomains: part–whole understanding, measurement interpretation, calculations, and word problems. As reflected in the Common Core State Standards (National Governors Association Center for Best Practices 2010), however, algebra is yet another major curricular strand at the elementary school level. This is a complicated curricular scope, and it is unclear whether strengthening performance in one domain can be expected to transfer to other components. Failure to produce strong performance across curricular components, as has been sometimes assessed and demonstrated, creates additional challenges for Tier 2 mathematics intervention (beyond those that are relevant for Tier 2 reading intervention) to circumvent the need for ongoing Tier 2 support in mathematics.

In this chapter, we focus is on the mathematics side of Tier 2 intervention. First, we provide an overview of the design principles involved in effective Tier 2 intervention and illustrate their application in a validated tutoring program for addressing students' difficulty with word problems. Then, we discuss more recent innovations in Tier 2 intervention by focusing on early arithmetic skill at first grade and on conceptual understanding and procedural skill

L. S. Fuchs (✉) · D. Fuchs · A. S. Malone
Vanderbilt University, Nashville, TN, USA
e-mail: lynn.fuchs@vanderbilt.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_18

with fractions at fourth grade. The chapter concludes with a discussion of the limitations of Tier 2 intervention research in mathematics and identifies areas for future research.

Principles of Effective Tier 2 Intervention

Six Design Principles

As conceptualized elsewhere (Fuchs et al. 2008), six design principles are central to the provision of effective Tier 2 intervention in mathematics. The first is *instructional explicitness*. Many typically developing students profit from the general education mathematics program that typically relies, at least in part, on a constructivist, inductive approach to instruction, in which teachers avoid explicitly explaining the nature of concepts or procedures for solving mathematics problems. Instead, teachers encourage students to discover conceptual understanding and methods for problem solution, with the hope that this will lead to deeper understanding and longer retention. Students who accrue serious mathematics deficits, however, fail to profit from such an approach in ways that result in understanding of the structure, meaning, and operational requirements of mathematics. A meta-analysis of 58 mathematics studies (Kroesbergen and Van Luit 2003) revealed that students with mathematics difficulty benefited more from explicit instruction than from discovery-oriented methods. Therefore, effective intervention for students with mathematics requires explicit, didactic instruction in which the teacher provides detailed explanations.

Explicitness is important, but it is not sufficient. A second and often overlooked principle of effective mathematics intervention is *instructional design that minimizes the learning challenge*. The goal is to anticipate and eliminate misunderstandings with precise explanations and with the use of carefully sequenced instruction so that the achievement gap can be closed as quickly as possible. This is especially important given the ever-changing and multiple demands of the mathematics curriculum.

The third principle of effective mathematics intervention is the requirement that instruction provide a *strong conceptual basis* for procedures. Supplemental intervention has a history of emphasizing *systematic practice,* a critical and fourth principle of effective practice. Yet, supplemental intervention has sometimes neglected the conceptual foundation of mathematics, and such neglect can cause confusion, learning gaps, and a failure to maintain and integrate previously learned content. In terms of systematic practice, note that this practice needs to be rich in *cumulative review,* the fifth principle of effective intervention.

The sixth principle concerns the need to incorporate *motivators to encourage students to work hard and regulate their attention and behavior*. Students with learning difficulty often display attention, motivation, and self-regulation difficulties, which may adversely affect their behavior and learning (e.g., Fuchs et al. 2005, 2006). By the time students enter Tier 2 intervention, they have experienced failure, causing many to avoid the emotional stress associated with learning mathematics. They no longer try for fear of failing. Therefore, intervention must incorporate systematic self-regulation strategies and motivators; for many students, tangible reinforcers are required. See Table 1 for a summary of the six design principles along with examples from each of the three Tier 2 mathematics interventions discussed below.

Illustrating the Six Design Principles with Pirate Math To illustrate these six design principles for effective Tier 2 intervention, we describe a validated tutoring program called Pirate Math (e.g., Fuchs et al. 2009). Pirate Math is designed to address arithmetic as well as word-problem difficulty, while building procedural calculation and early algebraic knowledge. We incorporated a pirate theme because in this schema-broadening instructional program, students are taught to represent the underlying structure of word-problem types using algebraic equations. “They find X, just like Pirates find X on treasure maps.”

Table 1 Implications for practice

Six principles of effective intervention	Examples from Pirate Math (grades 2–3)	Examples from Galaxy Math (grade 1)	Ex3amples from fraction challenge (grade 4)
Instruction is systematic and explicit	Scaffolded instruction on underlying structure of three word-problem types and solution strategies; all lessons include modeling, guided practice, independent practice, and corrective feedback; simple language is used for explanations, with frequent checks on understanding	Scaffolded instruction on arithmetic principles with manipulatives, number lines, and paper and pencil activities; all lessons include modeling, guided practice, independent practice, and corrective feedback; simple language is used for explanations, with frequent checks on understanding	Scaffolded instruction on fraction magnitude with manipulatives, number lines, and paper and pencil activities; all lessons include modeling, guided practice, independent practice, and corrective feedback; simple language is used for explanations, with frequent checks on understanding
Instructional design minimizes the learning challenge	Task analysis provides the most efficient procedures for teaching efficient problem-solving strategies	Task analysis provides the most efficient counting methods, taking into account first graders' fine motor skills	Task analysis provides the basis for instructional strategies for chunking and segmenting complex fraction comparison tasks
Instruction has a strong conceptual basis, while providing students with efficient procedural strategies	Instruction on each word-problem type begins with real-life examples and pictorial aids to help students understand the underlying structure of the problem type, while teaching students the most efficient approach word-problem solution strategies	Instruction dominantly focuses on the number knowledge foundational to arithmetic, while teaching students efficient counting strategies to solve addition and subtraction problems	Instruction focuses dominantly on the measurement interpretation of fractions, while teaching students procedural strategies (e.g., "bigger denominator, bigger fraction," and "fewer parts, bigger fraction")
Instruction includes systematic practice	In each lesson, "warm-ups" provide flash card practice for applying counting strategies to solve arithmetic problems; daily practice in solving word problems	In each lesson, speeded practice is provided on basic addition and subtraction facts	In each lesson, the Speed Game provides practice on fraction concepts (e.g., circling fractions equivalent to ½)
Instruction includes cumulative review.	All independent practice activities include both new material and review	All independent practice activities include both new material and review	All independent practice activities include both new material and review
Instruction includes motivators to reinforce positive behavior	Timer is used to monitor and award "gold coins" for on-task behavior on an interval basis; bonus problems (announced at the end of lessons) earn students coins; students spend coins on prizes	Stickers are provided for on-task behavior and hard work, which are affixed to "Galaxy Math Sticker Chart"; when chart is completed, students select a prize	Timer is used to monitor and award half or quarter "dollars" for on-task behavior; bonus problems (announced at end of lessons) earn students half or quarter dollars; students spend fraction money at "Fraction Store."

Pirate Math comprises four units: an introductory unit, which addresses mathematics skills foundational to solving word problems, and three word-problem units, each focused on a different type of word problem. Every tutoring lesson is scripted. Scripts are studied; they are not read or memorized. Pirate Math runs for 16 weeks, with 48 sessions (three per week).

Each session lasts 20–30 min. The instruction, as outlined below, is *systematic and explicit*; it is designed with care to *minimize the learning challenge*; it is rich in *concepts*; it incorporates *systematic practice* as well as *cumulative review*; and it relies on *systematic reinforcement* to encourage good attention, hard work, and accurate performance.

The introductory unit addresses mathematics skills foundational to word problems. Tutors teach strategic counting for deriving answers to arithmetic problems, review algorithms for double-digit addition and subtraction procedural calculations, teach methods to solve for "X" in any position in simple algebraic equations (i.e., $a+b=c$; $d-e=f$), and teach strategies for checking work in word problems.

A single strategic counting lesson is designed to address arithmetic deficits. Students are taught that if they "just know" the answer to an arithmetic problem, to "pull it out of your head." If, however, they do not know an answer immediately, they "count up." Strategic counting for addition and subtraction is introduced with the number line. *The "min strategy" is a mathematics counting strategy in which students start with the larger number and count up and the answer to the addition problem is the last number spoken. This is a more efficient counting strategy than the "max strategy" in which students start with the first number in an addition problem and count up (regardless of whether the first addend is the bigger or smaller number). The "min strategy" is referred to across the mathematics literature and is defined in the chapter.*

Practice in strategic counting is then incorporated into subsequent lessons. The tutor begins each session by asking the student, "What are the two ways to find an answer to a simple math problem?" The student responds, "Know it or count up." Then, the student explains how to count up an addition problem and how to count up a subtraction problem. Next, the tutor requires the student to count up two addition and two subtraction problems. Then the tutor conducts a flash card warm-up activity, in which students have 1 min to answer arithmetic problems. If they respond incorrectly, the tutor requires them to count up until they derive the correct answer. At the end of 1 min, the tutor counts the cards, and the student then has another minute to beat the first score. Also, throughout the lesson, whenever the student makes an arithmetic error, the tutor requires the student to count up. Finally, when checking the paper–pencil review, the tutor corrects arithmetic errors by demonstrating the counting strategy.

Each of the three word-problem units focuses on one word-problem type and, after the first problem-type unit, subsequent units provide systematic, mixed cumulative review that includes previously taught problem types. The word-problem types are Total (two or more amounts are combined; e.g., Doris has two flowers. Her sister has five. How many flowers do they have?), Difference (two amounts are compared; e.g., Doris is 2 years old. Her sister is 5. How much older is her sister?), and Change (initial amount increases or decreases; e.g., Doris had two pennies. Then she got two more. How much money does she have now?). Each word-problem session comprises six activities. The first is the counting strategies review and flash card warm-up already described, which lasts 5 min. Word-problem warm-up is the next activity, which lasts about 2 min and begins during the first word-problem unit. The tutor shows the student the word problem that the student had solved during the previous day's paper-and-pencil review. The student explains to the tutor how he or she solved the problem.

Conceptual and strategic instruction is the next activity. It is the heart of the lesson, lasting 15–20 min. Tutors provide scaffolded instruction in the underlying structure of and in solving the three types of word problems (i.e., developing a schema for each problem type), along with instruction on identifying and integrating transfer features (to broaden students' schema for each problem type). The tutor relies on role-playing, manipulatives, instructional posters, modeling, and guided practice. In each lesson, students solve three word problems, with decreasing amounts of support from the tutor.

Total is the first problem type addressed. In the total unit, tutors teach students to *run* through a problem: a three-step strategy prompting students to read the problem, underline the question, and name the problem type. Students use the *run* strategy across all three problem types. Next, for each problem type (i.e., schema), students are taught an algebraic equation to represent the underlying structure of that problem type. Students fill in slots of the equation as they identify and circle relevant information in the problem narrative. For example, for Total problems, students circle the item (e.g., flowers) being combined and the numerical values representing that item (e.g., 2 and 5), and then label the circled numerical values as "$P1$" (for part one; e.g., 2), "$P2$" (for part two; e.g., 5), and "T" (for the total; e.g., indicated with x—the missing value). Students then construct an equation representing the underlying mathematical structure of the problem type. For Total problems, the equation takes the form of "$P1+P2=T$," and "x" can appear in any of the three variable positions. Students are taught to solve for x, to provide a word label for the answer, and to check the reasonableness and accuracy of their work.

The strategy for Difference problems and Change problems follows similar steps but uses variables and equations specific to those problem types. For Difference problems, students are taught to look for the bigger amount (labeled "B"), the smaller amount (labeled "s"), and the difference between amounts (labeled "D"), and to use the algebraic equation "$\mathrm{B}-\mathrm{s}=\mathrm{D}$" to represent the problem type. For Change problems, students are taught to locate the starting amount (labeled "St"), the changed amount (labeled "C"), and the ending amount (labeled "E"); the algebraic equation for Change problems is "$\mathrm{St}\pm\mathrm{C}=\mathrm{E}$" ($\pm$ depends on whether the change is an increase or decrease in amount).

For each problem type, explicit instruction to broaden schemas occurs in six ways. First, students are taught that because not all numerical values in word problems are relevant for finding solutions, they should identify and cross out irrelevant information as they identify the problem type. Second, students learn to recognize and solve word problems with the missing information not only in the traditional, third slot of the equation, but also in the first or second position of the algebraic equation representing the underlying structure of the problem type. Third, students learn to apply the problem-solving strategies to word problems that involve addition and subtraction with double-digit numbers with and without regrouping. Fourth, students learn to solve problems involving money. Fifth, students are taught to find relevant information for solving word problems in pictographs, bar charts, and pictures. Finally, students learn to solve two-step problems that involve two problems of the same problem type or that combine problem types. Across the three problem-type units, previously taught problem types are included for review and practice.

Sorting word problems is the third activity and takes 5 min. Tutors read aloud flash cards, each displaying a word problem. The student identifies the word-problem type, placing the card on a mat with four boxes labeled "Total," "Difference," "Change," or "?." Students do not solve word problems; they sort them by problem type. To discourage students from associating a cover story with a problem type, the cards use similar cover stories with varied numbers, actions, and placement of missing information. After 2 min, the tutor notes the number of correctly sorted cards and provides corrective feedback for up to three errors.

In paper-and-pencil review, the final activity, students have 2 min to complete nine number sentences asking the student to find x. Then, they have 2 min to complete one word problem. Tutors provide corrective feedback and note the number of correct problems on the paper. Tutors require students to count up arithmetic errors, and keep the paper-and-pencil review sheet for the next day's word-problem warm-up activity.

Pirate Math includes a systematic reinforcement program. Throughout each Pirate Math session, tutors use a timer, which is set to ring three times at unpredictable intervals. If all students in the group are "on task" when the timer rings, each student earns a gold coin. (If one or more students are not on task, no one earns a coin.) Students can also earn gold coins for completing bonus problems correctly. Students do not know which problems are bonus problems until the end of the lesson. This encourages hard work throughout the session. At the end of the lesson, each earned gold coin is placed on the student's individual "treasure map." Sixteen coins lead to a picture of a treasure box and, when reached, the student chooses a small prize from a treasure box. The student keeps the old treasure map and receives a new map in the next lesson.

Effectiveness of Pirate Math The efficacy of Pirate Math has been demonstrated and replicated (e.g., Fuchs et al. 2009, 2010a). For example, Fuchs et al. (2009) identified third-grade students with substantial difficulty in computation and word problems in the Nashville-Metropolitan Public Schools and in the Houston Independent School District. Then, these children were randomly assigned to receive 13 weeks (3 times per week) of Pirate Math tutoring or Math Flash (Fuchs et al. 2003) tutoring (a validated program focused entirely on computation) or control (the school program without any research-based mathematics tutoring). Each session was audiotaped. A representative sample of lessons was coded for fidelity against the tutoring scripts, and fidelity was strong for both tutoring interventions. Students were pre- and posttested on computation and word-problem measures.

Pirate Math and Math Flash students improved comparably on computation and significantly more than students in the control group. The effect size (ES) comparing Math Flash to the control group was large (0.85 standard deviations). The ES comparing Pirate Math to the control group was similar, but somewhat smaller (0.72). But given that Pirate Math allocated only 5 min of every session to computation (whereas Math Flash spent 20–30 min per session on computation), Pirate Math's effects on computation are noteworthy. At the same time, however, effects on word problems clearly favored Pirate Math. The ES comparing Pirate Math to the control group was large (0.89), and there was no significant difference between Math Flash and the control group. Most impressively, the ES comparing Pirate Math to Math Flash was 0.72. This indicates that when Pirate Math, a tutoring program designed to incorporate the six principles of effective Tier 2 intervention, is implemented as designed, it benefits at-risk (AR) students' word-problem learning in dramatic ways on computation and word problems.

Innovations in Tier 2 Intervention

This section provides examples of innovations in Tier 2 intervention, relying on some of our more recent research (while providing an overview of prior related intervention work). The first focus is on early arithmetic skill (i.e., adding and subtracting single-digit numbers) at first grade. Early arithmetic skill (at the start of first grade) predicts mathematics learning through the end of fifth grade (Geary 2011) and is an indicator of risk for long-term learning disabilities (Geary et al. 2012b). The NMAP (2008) concluded that early mastery of simple arithmetic is a critical step toward eventual mastery of high-school algebra, a gateway for later entry into mathematics-intensive fields. These factors point to the importance of early mathematics intervention and the need for a strong focus on arithmetic in that intervention. Yet, despite its foundational importance, few randomized control trials have been conducted with early arithmetic Tier 2 intervention as its focus and, prior to our most recent arithmetic study, none had sought to isolate the added value of an intervention component. After describing our innovation in Tier 2 intervention for early arithmetic difficulty, we move to fraction intervention at fourth grade. Half of middle and high school students in the USA are still not proficient with the ideas and procedures taught about fractions in the elementary grades (e.g., Behr et al. 1984; Hiebert and Wearne 1985; National Council of Teachers of Mathematics (NCTM) 2007; NMAP 2008; Ni 2001). Yet, competence with fractions is considered foundational for learning algebra, for success with more advanced mathematics, and for competing successfully in the American workforce (NMAP 2008; Geary et al. 2012b). For these reasons, NMAP (2008) recommended that high priority be assigned to improving performance on fractions, a theme reflected in the Common Core State Standards (National Governors Association Center for Best Practices 2010) (http://www.corestandards.org). Therefore, improving common fraction performance for fourth graders at risk for poor outcomes is a critical focus. Yet, prior to this study, no prior work on Tier 2 intervention at the elementary school level was identified.

Early Arithmetic Skill at First Grade

Development of Arithmetic Competence and Need for Early Intervention By the time children enter first grade, most have a rudimentary understanding of addition and subtraction and can count to solve these problems (Geary 1994). For addition, they typically count both addends. For subtraction, they represent the beginning quantity (the minuend) with objects and sequentially separate the number of objects to be subtracted (the subtrahend); then they count the remaining set (e.g., Groen and Resnick 1977). As understanding of cardinality and the counting sequence develops, children discover the number-after rule for adding with 1. They also understand the sum of 5 + 2 cannot be 6 but instead is two numbers beyond 5. In this way, children discover the efficiency of counting from the first addend and rely on more efficient counting procedures. For addition, the most efficient counting procedure involves the min strategy: starting with cardinal value of the larger addend and counting up the number of times equal to the smaller addend (e.g., 4 + 3 = "four: five, six, seven"). For subtraction, the most efficient strategy relies on the missing addend: starting with the subtrahend and counting to the minuend (e.g., 5 – 2 = "two: three, four, five"; the number of counts is the answer). Frequent use of efficient counting procedures reliably results in the correct association between problem and answer, which produces long-term memories (Fuson and Kwon 1992; Siegler and Robinson 1982; Siegler and Shrager 1984). This enables direct retrieval of answers, and the commutativity of addition facilitates retrieval of related addition problems (Rickard et al. 1994). Subtraction, which is not commutative, is more difficult, but can be facilitated by retrieval of related addition facts (e.g., 8 – 5 = 3, based on 5 + 3 = 8; LeFevre and Morris 1999), once children understand the inverse relation between addition and subtraction (Geary et al. 2008).

Students with mathematics learning disabilities show consistent delays in the adoption of efficient counting procedures, make more counting errors during their execution, and fail to make the shift toward memory-based retrieval (e.g., Geary et al. 2012a; Goldman et al. 1988). Most of these children eventually catch up to peers in skilled use of counting procedures, but difficulty with retrieval tends to persist (Geary et al. 2012b; Jordan et al. 2003). They retrieve fewer answers from memory and when they do retrieve answers, they commit more errors (e.g., Geary et al. 1991; Geary et al. 2007). So, simple arithmetic fluency may be a signature deficit of mathematics learning disability (e.g., Geary et al. 2012b; Goldman et al. 1988; Jordan et al. 2003), and remediating arithmetic deficits in older students can be difficult (Fuchs et al. 2010a, c). For these reasons, there is a pressing need for early Tier 2 intervention.

Prior Work and Purpose of Our Innovation Four randomized control trials assessing Tier 2 intervention efficacy for first-grade students at risk for poor mathematics outcomes were located. Fuchs et al. (2006) conducted a randomized control trial to assess the efficacy of practice alone. An addition or subtraction problem with its answer briefly flashed on a computer screen; then students generated the problem and answer from short-term memory. This is based on the assumption that practice strengthens retrieval when problems and answers are simultaneously active in working memory (Geary 1993). Compared to an analogous computer-assisted spelling practice condition, arithmetic practice (10 min, twice weekly for 18 weeks) produced significantly better performance for addition but not subtraction; ESs were 0.95 and −0.01. Other mathematics outcomes were not assessed.

Two randomized control trials combined number knowledge tutoring with practice and assessed a broader range of outcomes. In Fuchs et al. (2005), tutoring occurred three times per week for 16 weeks. Each session included 30 min of tutor-led instruction designed to build number knowledge plus 10 min of computerized arithmetic practice, as just described. Results favored tutoring over a no-tutoring control group on measures of concepts and applications (ES = 0.67), procedural calculations (ES = 0.40-0.57), and word problems (ES = 0.48), but effects were not reliable on simple arithmetic (ESs = 0.15–0.40). Bryant et al. (2011) also integrated tutoring on number knowledge with practice (four times per week for 19 weeks). In each session, 20 min were devoted to number knowledge; 4 to practice, which focused on arithmetic problems, as well as reading numerals, counting on/back, writing dictated numerals, and writing 3-number sequences. Effects were significantly stronger for tutoring compared to a no-tutoring control group on simple arithmetic (ES = 0.55), place value (ES = 0.39), and number sequences (ES = 0.47). But tutoring did not enhance word-problem outcomes (ES = −0.05 and 0.07).

In the only randomized control trial to focus exclusively on number knowledge, Smith et al. (2013) evaluated Math Recovery (MR, Wright et al. 2002; Wright 2003), in which tutors adapt lessons to meet student needs as reflected on MR assessments. Tutors introduce tasks and have students explain their reasoning, but practice is not provided. Tutoring was to occur 4–5 times per week, 30 min per session across 12 weeks, but the median number of sessions was 32. At end of first grade, effects favored MR over the control group on fluency with simple arithmetic (ES = 0.15), concepts and applications (ES = 0.28), quantitative concepts (ES = 0.24), and mathematical reasoning (ES = 0.30). Effects were stronger for students who began tutoring below the 25th percentile (0.31–0.40), but are generally smaller than in studies that combined number knowledge tutoring with practice. Comparisons are, however, difficult because this study allowed fidelity to vary, whereas the other studies tried to ensure fidelity.

These four studies suggest potential for Tier 2 intervention, compared to no tutoring, for enhancing some forms of mathematics learning among AR first graders. Prior work does not, however, provide the basis for understanding whether the effects of Tier 2 intervention are simply a matter of more instruction (i.e., prior studies only include no-tutoring control groups rather than incorporating a contrasting tutoring condition). In this way, prior work also fails to inform practitioners about what components of Tier 2 intervention contribute to positive effects.

For these reasons, our innovation focused on the effects of number knowledge tutoring with contrasting forms of practice on AR first graders' emerging competence with simple arithmetic (Fuchs et al. 2013a). Two-digit calculation, number knowledge, and word-problem outcomes were also assessed. The major emphasis in tutoring was developing interconnected knowledge of number, but a small portion of each session was devoted to practice. In one condition, practice was designed to reinforce the relations and principles that serve as the basis of reasoning strategies that support fact retrieval. The other form of practice was more rote: It was designed to promote quick responding and use of efficient counting procedures to generate many correct responses and thereby form long-term representations to support retrieval. Both practice conditions occurred on the same content, encouraged strategic behavior, and provided immediate corrective feedback. So the two major distinctions between the two forms of practice were as follows. First, one condition encouraged a variety of number-principle strategies (e.g., relying on number lists, arithmetic principles such as cardinality, commutative principle, subtraction as the inverse of addition, and efficient counting procedures), whereas the other condition only encouraged efficient counting strategies. Second, in the condition that encouraged a variety of strategic behavior, practice did not involve speeded execution of the chosen strategy; the focus instead was on executing strategies thoughtfully to emphasize number knowledge. By contrast, in the condition that relied exclusively on counting strategies, practice was speeded. In this chapter, the terms *nonspeeded practice* and *speeded practice* are used to refer to these conditions.

To understand the efficacy of number knowledge tutoring when combined with speeded versus nonspeeded practice, we compared each tutoring condition against an AR no-tutoring control group that received the same classroom instruction as the tutored groups. To understand how the type of practice affects learning and whether the effects of tutoring are attributable to more than simply providing extra instructional time, we contrasted the two tutoring conditions against each other. To provide insight into whether different forms of tutoring help narrow the achievement gap, we included a group of low-risk classmates, who received the same classroom instruction as the AR tutored and control groups. We recruited forty schools and two hundred and thirty-three classes, in which we screened students to identify students with and without risk. Attrition was minimal, with 190 number knowledge tutoring + nonspeeded tutoring, 195 number knowledge tutoring + speeded practice, 206 AR no-tutoring control, and 300 low-risk control completing the study.

Nature of Tutoring Tutoring occurred for 16 weeks, three times per week, 30 min per session. In both tutoring conditions, 25 of each 30-min session were the same, designed to foster number knowledge. The last 5 min, which involved practice, differed. To foster engagement, the program uses a space theme (e.g., children are encouraged to "blast off into the math galaxy" by improving their mathematics knowledge; some manipulatives are shaped as space rockets). Tutors/students refer to the program as Galaxy Math (Fuchs et al. 2010b), which is related to Number Rockets (Fuchs et al. 2005).

Segment 1 focuses on number knowledge, with five units. Unit 1 (lessons 1–18) addresses basic number knowledge. Unit 2 (lessons 19–20) focuses on arithmetic doubles ($0+0$ through $6+6$; $2-2$ through $12-6$). Unit 3 (lessons 21–52) addresses arithmetic problem sets 5 through 12 (e.g., the 5 set includes all problems with sums or minuends of 5; the term *sets* is used, as in the *5 set*). Unit 4 (lessons 53–66) focuses on tens concepts. Unit 3 comprises approximately half the program (not counting the review unit). It focuses on partitioning number into constituent sets and number families (e.g., for the 5 set, $0+5$, $1+4$, $2+3$, $5-0$, $5-1$, $5-2$, etc.). Toward this end, five activities are conducted in each lesson. First, the tutor and student use unifix cubes to explore how the target number (e.g., 5 in the 5 set) can be partitioned in different ways to derive the adding and subtracting problems comprising that set. The second activity also focuses on part–whole knowledge, but with number families (problems using the same three numbers, e.g., $2+3=5$, $3+2=5$, $5-2=3$, and $5-3=2$, in that set). The tutor relies on visual displays that group families in the set and uses blocks to help the student rely on part–whole knowledge to understand how/why four problems make a family. Third, the student generates all addition and subtraction problems (with answers) in the target set, while using rockets to show the problems. Fourth, the tutor and student work together to solve a word problem on that set; they produce the answer and explain why the word problem is specific to the set. Fifth, the student reviews previous sets, orally stating answers to problems with corrective feedback. Between one and four lessons are allocated to each set, with a mastery test determining if and when students can advance before all four lessons on that are conducted.

Segment 2 involves practice. The content addressed in the final 5-min segment was the same in both conditions, addressing content covered in that day's number knowledge tutoring lesson. In the nonspeeded practice condition, students played games with space-themed manipulatives. Games were designed to provide contextualized review of the content addressed in the day's lesson. For example, to reinforce $n \pm 1$ lessons, the student used a spinner on a dial segmented from 1 to 19 to identify the number of "rockets called out to explore the math galaxy" and counted this number of rockets onto the board. Then, the tutor informed the student that either one more rocket was needed or that one rocket was called back to the space station, so the student added one rocket to the board or took one away. The student then stated the answer and number sentence. For an 8-set lesson, the student was informed how many rockets constituted the fleet and wrote that numeral as the total. The student then rolled a die to find the first group of rockets released from the space station, counted that number of rockets onto the game board, and wrote the numeral as an addend. The student then determined how many more rockets were needed to complete the fleet, wrote that numeral as an addend, and read the number sentence. Next, the student rolled the die to find how many rockets were called back to the space station. The student wrote numerals to generate and read a number sentence. Games differed for each day on the same topic. Throughout nonspeeded practice and lessons, tutors encouraged students to know the answer or to rely on a variety of number-principle strategies including (but not limited to) using number lists, relying on arithmetic principles (e.g., cardinality, commutative principle, inverse relation between addition and subtraction), and efficient counting strategies. "Knowing the answer right off the bat" was the preferred strategy when students were sure of answers.

In the speeded practice condition, students completed the "Meet or Beat Your Score" activity, with which students had 90 s to answer a stack of flash cards. For example, for $n \pm 1$ lessons, flashcards were all $n+1$ and $1+n$ problems ($n = 0$–18); for 8-set lessons, flashcards were all addition problems with the sum 8 and all subtraction problems with the minuend 8. Tutors corrected student errors immediately and required the student to use counting to produce the correct response. Therefore, students answered each problem presented correctly. The 90 s continued to elapse as the student used the counting procedure (as many times as needed). In this way, careful but quick responding increased the number of correct responses, which were counted and charted on a Rocket Chart at the end of 90 s. Then, the student had two chances to meet or beat that score. Throughout speeded practice and lessons, tutors required the student to know the answer (i.e., retrieve it from memory, if confident) or use the efficient counting strategies they had been taught. Tutors explained that "knowing the answer right off the bat" was preferred, if the child was sure of the answer. The counting strategies were simplified versions of the counting strategies already explained for Math Flash.

Findings In terms of fluency with simple arithmetic, number knowledge tutoring with nonspeeded practice produced significantly better learning compared to AR control students who did not receive tutoring. The ES was 0.38. Nonspeeded practice was designed to reinforce the relations and principles that were emphasized in the number knowledge portion of tutoring and that serve as the basis of reasoning strategies to support arithmetic skill. Findings lend theoretical support to studies indicating the important role number knowledge plays in developing competence with simple arithmetic (e.g., Baroody 1988, 1999; Butterworth and Reigosa 2007; De Smedt et al. 2009; Duncan et al. 2008; Koontz and Berch 1996; Rousselle and Noel 2007).

At the same time, nonspeeded practice did not help AR students narrow the achievement gap (ES = 0.07 favoring low-risk classmates). By contrast, incorporating speeded practice in number knowledge tutoring produced superior improvement in simple arithmetic compared to low-risk classmates, with an ES of 0.39, thereby narrowing the achievement gap. Speeded practice was also substantially more effective than number knowledge tutoring with nonspeeded practice (ES = 0.51) and produced an ES of 0.87 over AR control. In this way, this study extended earlier randomized control trials by isolating the effects of speeded practice, delivered in the context of tutoring to build number knowledge. Results indicate a substantial role for speeded practice in promoting simple arithmetic learning. It is noted that, in contrast to how practice is sometimes configured in schools (i.e., without sufficient scaffolding in number knowledge, in massed doses, and without support for correct responding), speeded practice was delivered in the context of number knowledge instruction and formulated practice to help children generate many correct responses, support development of fluency with efficient counting strategies, require students to immediately correct errors with an efficient counting procedure, and encourage strategic meta-cognitive behavior. So, findings should be generalized only to speeded practice that incorporates similarly sound, theoretically motivated instructional design.

Although findings suggest the value of speeded practice in supporting AR children's development of arithmetic competence, they do not address the possibility that speeded practice, with its focus on rote responding, is detrimental to other forms of mathematics learning. This possibility was investigated by assessing two-digit calculations, number knowledge, and word problems and found no evidence to support such a hypothesis. In fact, on complex calculations, speeded practice combined with number knowledge tutoring produced stronger learning compared to number knowledge tutoring with nonspeeded practice (ES = 0.21) or AR control (ES = 0.69). This was the case even though Unit 4's focus on multi-digit number knowledge was identical in the speeded and nonspeeded practice conditions. Moreover, improvement on complex calculations was comparable for speeded practice students and their low-risk classmates (ES = 0.01), even as improvement on more complex calculations for low-risk students exceeded that of nonspeeded practice students (ES = 0.19) and AR control students (ES = 0.57).

It is important to note, however, that effects on number knowledge or word-problem learning did not favor the speeded practice condition—although there was no indication that speeded practice was detrimental to these forms of mathematics learning. On number knowledge, the tutored groups developed comparably (ES = 0.11, favoring speeded practice) and better than AR control (ESs = 0.29 and 0.19 for the speeded and nonspeeded practice conditions, respectively). On word problems, the tutored groups again developed comparably (ES = 0.04, this time favoring nonspeeded practice) and better than AR control (ESs = 0.22 and 0.27 for speeded and nonspeeded practice, respectively).

In sum, findings suggest that number knowledge tutoring, with nonspeeded or speeded practice, is effective for enhancing arithmetic, complex calculations, number knowledge, and word-problem learning over no tutoring. At the same time, well-designed speeded practice, delivered in the context of tutoring to build number knowledge, is more effective than nonspeeded practice in promoting complex calculations as well as simple arithmetic, a core mathematical competence. Effects favoring speeded over nonspeeded practice on simple arithmetic (and complex calculations) were substantial, and results showed that the advantage for speeded over nonspeeded practice may occur by helping students compensate for the demands on reasoning ability, which an instructional focus on number knowledge creates. At the same time, no evidence that speeded practice inhibits development of number knowledge or word-problem skill was found, despite that rote responding was involved in speeded practice. In fact, both number knowledge tutoring conditions produced comparable number knowledge and word-problem learning, which was superior to AR control students (Fuchs et al. 2013a).

Conceptual Understanding and Procedural Skill with Fractions at Fourth Grade

Importance of Conceptual Understanding of Fractions Conceptual understanding is important for learning and maintaining accurate procedures with fractions (e.g., Byrnes and Wasik

1991; Hecht et al. 2003; Mazzocco and Devlin 2008; Ni and Zhou 2005; Rittle-Johnson et al. 2001). In this next study, we assessed the efficacy of a small-group tutoring program designed to foster understanding of fractions (Fuchs et al. 2013b). The intervention focused on two types of conceptual knowledge (Kieren 1993). The first is part–whole understanding, with which a fraction is understood as a part of one entire object or a subset of a group of objects. This type of understanding is typically represented using an area model, in which a region of a shape is shaded or a subset of objects is distinguished from the remaining objects. The second type of conceptual knowledge, the measurement interpretation of fractions, reflects cardinal size (Hecht 1998; Hecht et al. 2003) and is often represented with number lines (e.g., Siegler et al. 2011). In American schools, fractions are taught primarily via area models that underpin part–whole understanding. Measurement understanding is assigned a subordinate role (addressed later and with less emphasis).

Our study was innovative because our major emphasis was the measurement interpretation of fractions, although a smaller amount of time on part–whole conceptualizations to build on students' incoming understanding of fractions was also incorporated, as addressed in their classrooms. In emphasizing the measurement interpretation, we sought to avoid the understanding of fractions exclusively as part–whole relationships, which may create difficulty for conceptualizing improper and negative fractions (NMAP 2008). A focus on the measurement model is also in keeping with NCTM standards (2006) that instruction be designed to foster understanding of fraction magnitudes in terms of the number line. It is also an explicit emphasis of the fourth-grade Common Core State Standards on understanding of fraction equivalence and ordering (http://www.corestandards.org).

Prior Work and Purpose of Our Innovation Six studies assessing Tier 2 intervention efficacy on fractions were located. Bottge and colleagues conducted two studies in which they contrasted video-based real-world problem-solving instruction against conventional word-problem instruction. Both conditions required conceptual and procedural knowledge of fractions to solve problems. However, instruction on fractions, which occurred only in the service of problem-solving, was inductive and incidental. Bottge (1999) randomly assigned 17 eighth-grade students in a remedial mathematics class to the two ten-session conditions. Bottge et al. (2002) randomly assigned 42 seventh graders in two mathematics classes to 12 sessions and separately reported data for the eight students with prior mathematics difficulty. In both studies, there was little evidence of improvement on the fraction outcome, which required calculations, and no significant difference between conditions.

Other studies incorporated a more explicit instructional approach, reflecting the six design principles outlined earlier in this chapter. Relying on a multiple-baseline design, Joseph and Hunter (2001) demonstrated experimental control for a cue-card strategy across three eighth-grade AR students. A teacher initially taught students to use the cue card, which supported a three-pronged strategy for adding or multiplying fractions with and without common denominators and for reducing answers. After students showed competence in using the strategy, they were instructed to use the cue card while solving problems on daily fraction probes. In a maintenance phase, the cue card was removed. All three students showed substantial improvement with introduction of the cue card strategy, and maintenance (i.e., performance without the cue card) was strong. The study focus was, however, entirely procedural in terms of instruction and outcome.

Kelly et al. (1990) also took an explicit approach to fraction instruction, but focused simultaneously on procedures and concepts. They randomly assigned 28 high-school AR students from one remedial and one general mathematics class to ten sessions of teacher-mediated videodisc-supported instruction or teacher-mediated conventional textbook instruction. Although direct instruction was employed in both conditions, videodisc instruction differed by providing mixed problem-type instruction, separating highly confusable concepts and terminology during early instructional stages, and providing a broader range of examples to avoid misconceptions (e.g., introducing proper and improper fractions in the first lesson). Both groups improved substantially from pretest (40% on a 12-item test) to posttest (96% vs. 82%), with the videodisc group improving significantly more. Yet, despite the instructional focus on concepts and procedures, the fraction measure was largely procedural.

By contrast, intervention in the next two studies focused primarily on understanding of fractions and assessed outcomes on concepts as well as procedures. Butler et al. (2003) contrasted two explicit instruction conditions with 50 sixth- through eighth-grade AR learners. Both conditions carefully transitioned students from a conceptual emphasis, largely based on part–whole understanding, to algorithmic rules for handling fractions, and from visual to symbolic representations. Only one condition, however, included concrete manipulatives. Both groups significantly improved across ten sessions. On one measure, in which students circled fractional parts of sets, those who received 3 days of manipulatives improved significantly more; on the other four measures, the difference between conditions was not significant, providing mixed evidence regarding the importance of concrete representations within an explicit instructional approach. Without random assignment or a control group, however, conclusions are tentative.

Hecht (2011) expanded on Butler et al. (2003) by employing random assignment, including a control group, and doubling the duration of intervention. Seventh-grade AR students ($n=43$) were randomly assigned to control or 1:1 tutoring using 23 Rational Number Project lessons (Initial Fraction Ideas; Cramer et al. 2009). These lessons rely on area models to teach part–whole relations, the concept of the unit, order and equivalence, and addition/subtraction, and apply these concepts in the context of computation, word problems, and estimation of sums/differences. Intervention students improved significantly more on a range of procedural and conceptual measures, including a number line task rooted in measurement understanding of fractions, despite that none of the lessons addressed the measurement model or strategies for using the number line.

These studies provide the basis for only tentatively concluding that Tier 2 intervention, based on part–whole understanding of fractions, enhances fraction learning among middle- and high-school AR students. Each of these studies was, however, small and relied exclusively on experimental measures, closely aligned with instruction, to assess outcomes (an exception is Butler et al. 2003, who included a commercial criterion-referenced measure in addition to experimental tasks). More importantly, none of these studies addressed the earlier grades, when the foundation for understanding fractions is developed, and none operationalized risk with the clarity needed to understand the level of students' mathematics performance.

Our major innovation, therefore, was to examine the effects of early intervention, with the clarity required to understand what "risk" entailed (Fuchs et al. 2013b). We focused on fourth grade, when the curriculum emphasizes understanding of fractions. Conceptual understanding and calculation skill were assessed using experimental tasks as well as an external, widely accepted measure of competence with fractions: the pool of easy, medium, and hard fourth-grade and easy eighth-grade released fraction items from the National Assessment of Educational Progress (NAEP; 18 items were selected from the pool of items released between 1990 and 2009).

In Fuchs et al. (2013b), we operationalized risk as whole-number calculation skill below the 35th percentile. Whole-number skill was used to define risk because the range of performance on fraction measures is limited at the start of fourth grade and because prior work (Hecht and Vagi 2010; Seethaler et al. 2011) identifies whole-number calculation skill as a predictor of fraction learning. A cutoff at the 35th percentile is, however, high. So, we designed the study to examine whether response to intervention differs for students with more versus less severe incoming deficits with whole numbers (< 17th vs. 18th–34th percentile). This question, whether intervention is differentially effective depending on the severity of mathematics difficulty, is important for designing instruction that addresses the full range of students with mathematics difficulty. Yet, the authors were unable to locate experimental studies on this topic. Finally, to contextualize results, we compared year-end performance for the more and less severe groups of tutored students against low-risk (> 34th percentile) classmates. This is important given that few fraction measures provide a normative framework or thorough behavior sampling of fourth-grade fraction skill.

Nature of Tutoring The fraction intervention program, Fraction Challenge (Fuchs and Schumacher 2010), described in Fuchs et al. (2013b), includes 36 lessons taught over a 12-week period (three 30-min lessons per week). As with Galaxy Math, scripts provide a model of the lessons and key explanatory language. Tutors review scripts prior to delivering lessons; however, to promote teaching authenticity and responsiveness to student difficulty, tutors do not memorize or read scripts.

As mentioned, tutoring relies primarily on the measurement conceptualization of fractions, with content focused primarily on representing, comparing, ordering, and placing fractions on a 0–1 number line. This focus is supplemented by attention to part–whole understanding (e.g., showing objects with shaded regions) and fair shares representations to build on classroom instruction. In this way, number lines, fraction tiles, and fraction circles are used throughout the lessons, with stronger emphasis on part–whole representations in beginning lessons. Lesson 22 (of 36) introduces fraction computation. Throughout the program, we focus on proper fractions and fractions equal to one. Improper fractions greater than 1 are introduced with addition and subtraction of fractions. To reduce computational demands, denominators do not exceed 12 and exclude 7, 9, and 11. This tutoring content mirrored classroom instruction with the following exceptions. The focus was narrower, with greater emphasis on measurement understanding, whereas classroom instruction emphasized part–whole understanding more than intervention; calculations substantially less than classroom instruction were focused; and a more limited pool of denominators was used.

More specifically, during the first 2 weeks of Fraction Challenge, the focus is on understanding of fraction magnitude. We begin by addressing "what is a fraction" and teach relevant vocabulary (e.g., *numerator*, *denominator, unit*). We rely on a combination of part/whole relations, measurement, and equal sharing to explain the fraction magnitudes. Instruction emphasizes the role of the numerator and denominator and how they work together to constitute the fraction, which is one number, even though it comprises two whole numerals.

In the 3rd week, tutors review material presented in the first six lessons. Students practice naming fractions, reading fractions, and comparing two fractions when the denominators are the same or when the numerators are the same. In this review, two types of flashcards are used to build fluency with the meaning of fractions. The first type shows flashcards with one fraction; students read and state the meaning of the fraction. For example, for $\frac{1}{4}$ students say, "one-fourth, one of four equal parts." Students take turns over a 2-min period, responding to as many fractions as they can. The tutor keeps track of the group total for each lesson; the students' goal is to meet or beat the previous day's score. The second type shows two fractions. Students determine if the fractions pairs fit one of three categories: same numerators (different denominators), same denominators (different numerators), or different numerator and different denominator. Students categorize flashcards in this way for 1 min. Then, the tutor gives each student two fraction cards; for each, students place the greater than or less than sign between fractions and explain their rationale to the group.

In weeks 4 and 5, students learn about fractions equivalent to $\frac{1}{2}$ (e.g., $\frac{2}{4}, \frac{3}{6}, \frac{4}{8}, \frac{5}{10}, \frac{6}{12}$). They also learn to compare two fractions in which the numerators and denominators both differ, using $\frac{1}{2}$ as a benchmark for comparison and writing the greater than, less than, or equal sign between the fractions. Then, two activities are introduced: placing two fractions on the 0–1 number line, marked with $\frac{1}{2}$, and ordering three fractions from smallest to largest.

In week 6, tutors introduce fractions representing a collection of items and fractions equivalent to 1, while continuing to work on comparing two fractions, ordering three fractions, and placing fractions on the 0–1 number line, now without the $\frac{1}{2}$ marker. Students are encouraged, however, to think about where $\frac{1}{2}$ goes on the number line in relation to placing other fractions. Week 7 (lessons 19–21) was cumulative review on all concepts and skills.

Weeks 8 and 9 focus on simple calculations. Addition with like denominators is introduced first; then subtraction with like denominators; then mixed addition and subtraction; and then addition with unlike denominators and then subtraction with unlike denominators. Concepts and procedures are addressed. When introducing unlike denominators, tutors limit the pool of problems. In all cases, one fraction is equivalent to $\frac{1}{2}$ or 1 so students can write equivalent fractions they already learned. The last 2 weeks are cumulative review.

Each lesson comprises four activities: introduction of concepts or skills, group work, the speed game, and individual work. The first activity, introduction of new concepts and skills, lasts 8–12 min. Concrete manipulatives (e.g., fraction tiles, fraction circles), visual representations, and problem-solving strategies are presented. In group work, which lasts 8–12 min (the introduction plus group work lasted 20 min for each lesson), students rehearse and apply concepts and practiced strategies addressed in the introduction. Students take turns leading the group through problems, while all students show their work for each problem. The third activity, the speed game, is designed to build fluency on a previously taught concept or skill. For instance, to build fluency on fractions equivalent to $\frac{1}{2}$, tutors give each student a paper showing 25 fractions, and students have 1 min to circle fractions equivalent to $\frac{1}{2}$. Sometimes, the Speed Game requires computation, and students are given specific instructions on which items to solve. For example, students might be told to solve only addition problems or solve only problems with like denominators. In this way, students discriminate between problem types. The fourth activity is individual work for which students independently complete a two-sided practice sheet. One side presents problems taught in that day's lesson; the other side is cumulative review. This activity lasts approximately 8 min, for a total of 30 min per session.

As with Galaxy Math, the program also encourages students to regulate their attention/behavior and to work hard. Tutors teach students that on-task behavior means listening carefully, working hard, and following directions and that on-task behavior is important for learning. Tutors set a timer to beep at three unpredictable times during each lesson. If all students are on task when the timer beeps, all students receive a checkmark. To increase the likelihood of consistent on-task behavior, students cannot anticipate time intervals. Also, on each practice sheet, 2 of 16 problems are bonus problems. As the tutors score the practice sheet, they reveal which problems are bonus items. Students receive a checkmark for each correctly answered bonus problem. At the end of the lesson, tutors tally checkmarks for each student and award them with a "half dollar" per checkmark. At the end of each week, students shop at the "fractions store" to spend money earned during tutoring. All items in the store are listed in whole dollar amounts at three price points so students must exchange half dollars for whole dollars and determine what they can afford. In this way, to use the fraction store, students must rely on their fraction knowledge, while exercising judgment about buying a less expensive item versus saving for a more expensive one. In lesson 19, half dollars are replaced with quarter dollars.

Findings In our study, we stratified by risk severity and classroom when randomly assigning students to tutoring versus control conditions. Participants were fourth graders from 53 classrooms in 13 schools. Of these students, 259 were at risk: 129 tutored students (60 more severe and 69 less severe) and 130 control students (66 more severe and 64 less severe). The other 292 students were low-risk classmates. With this sample, we included two measures that isolated the type of understanding on which we primarily focused: measurement understanding. On comparing fractions (in which students place a greater than, less than, or equal sign between two fractions), the ES favoring AR intervention students over control children was 1.82 SDs, and the achievement gap between AR tutored students and their low-risk classmates narrowed, while the gap for AR control students increased. On fraction number line (in which students place a fraction on a 0–1 number line), the ES was 1.14. Although fraction number line data on low-risk classmates (only group-administered data on low-risk students were collected) were not collected, the posttest performance of AR tutored students was at the 75th percentile for a normative sample of sixth graders, as per Siegler et al. (2012).

Because the alignment for comparing fractions and fraction number line was greater for intervention than for classroom instruction, we also considered effects on NAEP. NAEP was not aligned with intervention, and it focused with comparable emphasis on measurement and part–whole understanding. Here, effects were also significant and strong. The ES favoring AR intervention students over control was 0.94 SDs, and the achievement gap favoring low-risk classmates over AR control students remained large, while the achievement gap for AR intervention students decreased substantially or was eliminated.

Moreover, although classroom instruction focused on calculations more than intervention, effects again favored intervention students over control. Here, the ES favoring AR intervention students was 2.51; the achievement gap between AR tutored students and their low-risk classmates narrowed, while the gap for AR control students increased; and AR tutored students' posttest performance actually exceeded that of low-risk classmates. Given that classroom instruction allocated substantially more time to calculations, this suggests that understanding fractions, perhaps specifically the measurement understanding of fractions, transfers to procedural skill, at least with respect to adding and subtracting fractions (Hecht et al. 2003; Mazzocco and Devlin 2008; Ni and Zhou 2005; Rittle-Johnson et al. 2001; Siegler et al. 2011).

In these ways, this study innovatively extends Tier 2 intervention on fractions by focusing primarily on understanding magnitude (rather than part–whole understanding as in earlier work) and targeting younger students (rather than middle or high school students). Another interesting extension to the literature concerns the focus on risk severity. We found that response to intervention was comparable for students with more versus less severe risk, when risk was defined in terms of whole-number deficits. That is, there were no significant interactions between risk severity and intervention condition, and ESs were similar for more versus less severe student groups. By any standard, the effects of intervention designed to foster understanding of fraction magnitude for AR fourth graders were strong, with the achievement gap for AR learners substantially narrowed or eliminated. Of course, classroom instruction's failure not only to address the needs of a substantial majority of AR learners in a more successful manner but also to promote stronger learning among low-risk classmates raises questions about the quality and nature of classroom fraction instruction. This in part explains widespread difficulty with fractions (e.g., Behr et al. 1984; NCTM 2007; Ni 2001) and highlights the pressing need to improve the quality of fraction instruction and learning in this country (NMAP 2008).

Need for Future Research

Studies described in this chapter, from our own research programs as well as prior work, provide the basis for thoughts about the power as well as the limitations of Tier 2 intervention. First, we focus on the power. As the research described in this chapter illustrates, the literature indicates it is possible to design tutoring programs to enhance the outcomes of students who are at risk of poor mathematics development. Interventions that incorporate explicit instruction, provide students with a strong conceptual foundation and efficient procedural strategies, and embed regular, strategic, and cumulative practice are generally efficacious. Results clearly demonstrate that students AR for poor mathematics development suffer reliably and substantially less positive mathematics outcomes if left in the general education program without such tutoring (as represented in the control conditions in these studies). When AR students do not receive these preventative tutoring services, the gap between their level of mathematics performance and those of low-risk classmates grows, making it increasingly difficult for these children to profit from classroom instruction. By contrast, AR students as a group, who receive high-quality tutoring, make progress toward catching up to classmates and, for some of these children, the scaffolding provided through such Tier-2 validated intervention creates a strong foundation for them to experience long-term success in mathematics. Clearly, reliable screening of risk to identify students for 12–20 weeks of accurately implemented, validated tutoring, as represented in Tier 2 of multilevel RTI prevention systems, is a valuable and important service.

At the same time, it is important for policymakers and schools to recognize the limitations of Tier 2 intervention for dramatically reducing the need for ongoing and intensive services for some segment of the school population. We focus on two such limitations: lack of universal response and questions about transfer across the components of the mathematics curriculum. In terms of lack of universal response, in the three studies from our own research program described in this chapter, not all students respond. This phenomenon has been documented before, not only for mathematics but also for reading As O'Connor and Fuchs (2013) described, for example, the modal rate of unresponsiveness on the components of the curriculum targeted for intervention approximated in our prior studies approximates 4 % of the general population. This is similar to the prevalence of learning disabilities in the USA when intelligence quotient (IQ)–achievement discrepancy is used as the method of identification (although research indicates that the groups of students identified via RTI methods of identification versus the IQ–achievement discrepancy differ; Fuchs et al. 2005a; Fuchs et al. 2008). This rate of unresponsiveness suggests the limitations of Tier 2 intervention for dramatically reducing the need for ongoing, intensive services for students traditionally identified as having a learning disability. This is the case when 12–20 weeks of small-group tutoring are provided. If it were possible to provide a longer duration of tutoring or deliver that tutoring individually, it may be possible to reduce the rate of unresponsiveness further. But with longer runs of one-to-one tutoring, services begin to resemble the level of intensity expected in special education, and this prompts concerns about due process and how schools might fund such a level of intensity without special education resources.

At the same time, it is important to consider that the rate of unresponsiveness in efficacy studies, which control the quality of implementation, probably underestimates the actual percentage when Tier 2 intervention is practiced in schools. In actual practice, it is likely that fidelity of implementation will be lower, with reduced effects. In addition, as students continue in school, the effects of tutoring can be expected to diminish, and without additional support, some responders will reemerge with difficulty.

The second issue that represents a challenge to preventing long-term mathematics difficulty with Tier 2 intervention concerns questions about transfer across components of the mathematics curriculum. As our studies described in this chapter illustrate, although transfer may occur across some domains it is decidedly limited across others. For example, in Fuchs et al. (2013a), we found clear indications of transfer from simple arithmetic tutoring to more complex calculations. However, transfer to word problems was more limited, with word-problem achievement gaps growing over the course of intervention for both tutoring conditions. In a similar way, in Fuchs et al. (2013b), although fraction intervention had dramatic effects in decreasing the achievement gap for more and less severe risk fourth graders, fraction tutoring had no effect on students' whole-number calculation skill. As already discussed, mathematics, more than reading, is potentially complicated by the fact that the elementary school curriculum comprises multiple components within and across the grades. This problem becomes more complicated over the course of high school, where the components of the mathematics curriculum (e.g., geometry, trigonometry, calculus, as well as algebra) diverge more dramatically than in the earlier grades.

Clearly, additional research on other components of the mathematics curriculum, at early and later stages of mathematics development, is required to elucidate where prior Tier 2 intervention creates protection against further risk and where one can expect new forms of risk to emerge. New risk may emerge due to lack of transfer from earlier intervention. Alternatively, new topics in the mathematics curriculum may create risk for students whose prior mathematics performance has been adequate. All this creates the need not only for additional intervention work, with a focus on long-term outcomes, but also for additional research on screening for risk on topics at the intermediate grades and at the middle- and high-school levels.

References

Baroody, A. J. (1988) Number comparison learning by children classified as mentally retarded. *American Journal on Mental Retardation, 92,* 461–471.

Baroody, A. J. (1999). The development of basic counting, number, and arithmetic knowledge among children classified as mentally handicapped. *International Review of Research in Mental Retardation, 22,* 51–103.

Behr, M. J., Wachsmuth, I., Post, T. R., & Lesh, R. (1984). Order and equivalence of rational numbers: A clinical teaching experiment. *Journal for Research in Mathematics Education, 15,* 323–341.

Bottge, B. A. (1999). Effects of contextualizing math instruction on problem solving of average and below-average achieving students. *Journal of Special Education, 33,* 81–92.

Bottge, B. A., Heinrichs, M., Mehta, Z., & Hung, Y. (2002). Weighing the benefits of anchored math instruction for students with disabilities in general education classes. *The Journal of Special Education, 35,* 186–200.

Bryant, D. P., Brian, B. R., Roberts, G., Vaughn, S. Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, 78, 7–23.

Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research and Practice, 18,* 99–111.

Butterworth, B., & Reigosa, V. (2007). Information processing deficits in dyscalculia. In B. Butterworth & V. Reigosa (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 65–81). Baltimore: Paul H Brooks Publishing.

Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural knowledge. *Developmental Psychology, 27,* 777–786.

Cramer, K., Behr, M., Post, T., & Lesh, R. (2009). *Rational number project: Initial fraction ideas*. Minneapolis: University of Minnesota College of Education and Human Development.

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numeral magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103,* 469–479.

Dirks, E., Spyer, G., van Lieshout, E. C. D. M., & de Sonneville, L. (2008). Prevalence of combined reading and arithmetic disabilities. *Journal of Learning Disabilities, 41,* 460–473.

Duncan, G. J., Dowsett, C. J., & Classens, A., et al. (2008). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Fuchs, L. S., & Schumacher, R. F. (2010). *Fraction Challenge*. Unpublished manual.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). Math Flash [computer software]. (Available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203).

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29–43.

Fuchs, L. S., Fuchs, D., Powell, S., Seethaler, P., Cirino, P., & Fletcher, J. (2008). Intensive intervention for students with mathematics difficulties: Seven principles for effective practice. *Learning Disability Quarterly*, 31, 79–92.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., Hamlett, C. L., & Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology, 101,* 561–576.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010a). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learning and Individual Differences, 20,* 89–100.

Fuchs, L. S., Fuchs, D., & Bryant, J. (2010b). *Galaxy Math*. Unpublished manual.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Fuchs, D., Hamlett, C. L., Cirino, P. T., & Fletcher, J. M. (2010c). A framework for remediating the number combination deficits of students with mathematics difficulty. *Exceptional Children, 76,* 135–156.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., Deselms, J., Seethaler, P. M., Wilson, J., Craddock, C. F., Bryant, J. D., Luther, K., & Changas, P. (2013a). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology, 105,* 58–77.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., & Changas, P. (2013b). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105,* 683–700.

Fuson, K. C., & Kwon, Y. (1992). Korean children's single digit addition and subtraction: Numbers structured by ten. *Journal for Research in Mathematics Education, 23,* 148–165.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345–362.

Geary, D. C. (1994). Developing arithmetic skills. In *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47,* 1539–1552.

Geary, D. C., Brown, S. C, & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27,* 787–797.

Geary, D. C., Hoard, M. K., Byrd-Craven, J, Nugent, L., & Numtee, C. (2007) Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1359.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33,* 277–299.

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012a). Fact Retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45,* 291–307.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012b). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104,* 204–223.

Goldman, S. R., Pellegrino, J. W., & Mertz, D. L. (1988). Extended practice of addition facts: Strategy changes in learning-disabled students. *Cognition and Instruction, 5,* 223–265.

Groen, G., & Resnick, L. B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology, 69,* 645–652.

Hecht, S. A. (1998). Toward an information-processing account of individual differences in fraction skills. *Journal of Educational Psychology, 90,* 545–559.

Hecht, S. A. (2011). *Effects of a conceptual knowledge intervention for seventh-graders with fraction difficulties*. Unpublished manuscript. Available from Steven A. Hecht, Nova Southeastern University,1750 NE 167th Street, North Miami Beach, FL 33162–3017.

Hecht, S. A., & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology, 102,* 843–859.

Hecht, S. A., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86,* 277–302.

Hiebert, J., & Wearne, D. (1985). A model of students' decimal computation procedures. *Cognition and Instruction, 2,* 175–205.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85,* 103–119.

Joseph, L. M., & Hunter, A. D. (2001). Differential application of a cue card strategy for solving fraction problems: Exploring instructional utility of the Cognitive Assessment System. *Child Study Journal, 31,* 123–136.

Kelly, B., Gersten, R., & Carnine, R. (1990). Student error patterns as a function of curriculum design: Teaching fractions to remedial high school students and high school students with learning disabilities. *Journal of Learning Disabilities, 23,* 23–29.

Kieren, T. E. (1993). Rational and fractional numbers: From quotient fields to recursive understanding. In C. P. Thomas, E. Fennema & R. A. Thomas (Eds.), *Studies in mathematical thinking and learning* (pp. 49–84). Hillsdale: Lawrence Erlbaum Associates.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition, 2,* 1–23.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special needs: A meta-analysis. *Remedial and Special Education, 24,* 97–114.

LeFevre, J., & Morris, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition, 27,* 803–812.

Mazzocco, M. M. M., & Devlin, K. T. (2008). Parts and 'holes': Gaps in rational number sense among children with vs. without mathematical learning disabilities. *Developmental Science, 11,* 681–691.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading, 9,* 85–116.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. Washington, DC: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2007). *The learning of mathematics*. Reston: Author.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common core state standards (Common Core State Standards for Mathematics). Washington, DC: Author. Retrieved from http://www.corestandards.org/

National Mathematics Advisory Panel. (2008). *The Final Report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Ni, Y. (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemporary Educational Psychology, 26,* 400–417.

Ni, Y., & Zhou, Y. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist, 40,* 27–52.

O'Connor, R., & Fuchs, L. S. (2013). Responsiveness to intervention in the elementary grades: Implications for early childhood education. In V. Buysse, E. Peisner-Feinberg, & J. Cantler (Eds.), *Handbook of response to intervention (RTI) in early childhood education*. Baltimore: Brookes.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1139–1153.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346–362.

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *The Journal of Human Resources, 27,* 313–328.

Rousselle, L., & Noel, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs. non-symbolic number magnitude. *Cognition, 102,* 361–395.

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21,* 536–542.

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child and Adolescent Pcyhology, 9,* S58–S64.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsett (Eds.), *Advances in child development and behavior* (pp. 242–312). New York: Academic Press.

Siegler, R. S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale: Erlbaum.

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62,* 273–296.

Siegler, R. S., Duncan, G .J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. Psychological Science, 23, 691–697.

Smith, T. B., Cobb, P., Farran, D., Cordray, D., & Munter, C. (2013). Evaluating math recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal, 50,* 397–428.

Wright, R. J. (2003). Mathematics Recovery: A program of intervention in early number. *Australian Journal of Learning Disabilities, 8,* 6–11.

Wright, R. J., Martland, J., Stafford, A. K., & Stanger, G. (2002). *Teaching Number: Advancing children's skills and strategies*. London: Paul Chapman Publishing.

Implementation of Tier 2 Reading Interventions in the Primary Grades

Jeanne Wanzek, Stephanie Al Otaiba and Brandy Gatlin

A substantial body of research has provided educators with essential guidance for implementing effective early intervention for students with reading difficulties (e.g., Blachman et al. 2004; Torgesen et al. 1999; Vaughn et al. 2003; Vellutino et al. 1996). Spawning out of this research, multi-tiered models of reading instruction and intervention were developed to provide coordinated efforts for the implementation of effective instruction to prevent and remediate reading difficulties in the earliest grades (Denton et al. 2006; McMaster et al. 2005; O'Connor et al. 2005; Vaughn et al. 2009). In these multi-tiered models, Tier 1 instruction provides the basis for reading instruction for all students, and at the early elementary level, is the core instruction for learning to read. Tier 2 interventions are then designed to supplement Tier 1 instruction for students who demonstrate risk or difficulties in their early attempts at learning to read. Students are identified for Tier 2 reading interventions through universal screening measures implemented at each grade level. At any time point, students who demonstrate difficulty meeting grade level reading expectations despite strong, effective, research-based Tier 1 reading instruction are identified for supplemental Tier 2 intervention.

Tier 2 interventions are intended to provide a more intense and targeted instructional environment to meet the needs of students who are at risk for or demonstrating reading difficulties, providing them the necessary instruction to accelerate learning and achieve grade level expectations. Tier 2 interventions are more intense and targeted than Tier 1 instruction both in content and delivery. For example, teachers must make specific decisions about the content to target in an intervention by identifying student strengths and weaknesses using screening and progress monitoring information. In addition, decisions must be made regarding the delivery of an intervention including the amount of time students will receive the intervention, the frequency of the intervention sessions, and the instructional group size needed to provide explicit, targeted instruction, practice, and feedback in the areas of student need. Thus, Tier 2 intervention is data-driven not only for identifying students in need of intervention but also in the planning and implementing of the intervention as well.

Research Base for Tier 2 Interventions

There is strong research evidence and a clear convergence of positive findings for providing Tier 2 interventions for students with reading

J. Wanzek (✉)
Florida State University, 1107 W. Call St.,
P.O. Box 306-4304, Tallahassee, FL 32306, USA
e-mail: jwanzek@fcrr.org

S. Al Otaiba
Southern Methodist University, Dallas, TX, USA

B. Gatlin
Florida State University, Tallahassee, FL, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_19

difficulties in the primary grades. In a systematic review of the relevant research, a panel of researchers (Gersten et al. 2008) noted several content and delivery aspects of Tier 2 implementation that were associated with improved decoding and comprehension outcomes for students with reading difficulties.

Content of Tier 2 Interventions

Gersten et al. (2008) recommended up to three foundational reading skills be taught based on student needs in the areas of comprehension, fluency, phonemic awareness, phonics, and vocabulary. That said, some components of instruction had a direct, casual research base and some of the components lacked direct research but were recommended by the panel until further research could assist in decision-making. For example, at the kindergarten level, phonemic awareness and letter sound instruction were supported directly by research studies. The panel also recommended listening comprehension and vocabulary development based on the developmental reading needs of students and the correlational research supporting the importance of these skills in successful reading. More recent studies have confirmed the potential of Tier 2 interventions for improving vocabulary knowledge at the kindergarten level (Luftus et al. 2010; Tuckwiller et al. 2010). At the Grade 1 level, inclusion of phonemic awareness, phonics, text fluency, vocabulary, and comprehension were supported directly by research studies, and the panel also recommended instruction and practice in word reading fluency to build automaticity in the first part of first grade when students may not be able to read enough text to adequately practice text fluency. In grade 2, the inclusion of phonics and vocabulary were directly supported by research, and the panel also recommended fluency with connected text and comprehension as critical components to Tier 2 intervention.

Along with these recommendations, the panel noted that fairly basic levels of comprehension instruction occurred in many of the studies with effective interventions, and only one study was located that implemented vocabulary instruction as a part of Tier 2 intervention (Gersten et al. 2008). Thus, the authors cautioned that it is possible inferential comprehension and vocabulary development may be better developed with more heterogeneous groups as a part of Tier 1 instruction. Recent work on Tier 2 vocabulary interventions has shown some promise for improving student vocabulary knowledge at kindergarten and first grade (Luftus et al. 2010; Pullen et al. 2010; Tuckwiller et al. 2010). In these studies, target words from Tier 1 instruction were reinforced in 20-min, small-group Tier 2 sessions provided twice a week to students demonstrating low vocabulary knowledge. The findings suggested students with low vocabulary knowledge may obtain greater knowledge of word meanings when words are taught both in Tier 1 classroom instruction and Tier 2 small-group instruction. However, the students did not seem to retain increased vocabulary knowledge levels over time, suggesting these supplemental vocabulary interventions may not have been intensive enough for long-term vocabulary development (Luftus et al. 2010; Pullen et al. 2010).

Regardless of the specific reading components targeted in an intervention, it is clear from all of the intervention studies reviewed by Gersten et al. (2008) that Tier 2 instruction in the targeted instructional components should be highly explicit and systematic, with clear explanations and teacher models provided for successfully accomplishing each new skill. Instruction and practice should systematically build from simple to more complex skills and concepts, allowing student mastery at each level with multiple opportunities for student practice and feedback.

Delivery of Tier 2 Interventions

Gersten et al. (2008) also made specific recommendations for delivery of Tier 2 interventions such as providing interventions in 20–40-min sessions provided three to five times per week in small, homogeneous groups of three to four students. Several studies with positive effects have implemented interventions with instruction provided either 1:1 or in small groups; however with no clear differences in effects noted between these studies, Gersten et al. recommended small groups because it could be considered more practical for implementation at the Tier 2 level. The amount of time dedicated to each intervention session as well as the frequency of sessions directly relate to the dosage of intervention students receive. Thus, session time and frequency of intervention sessions should be determined based on students' needs including the number of reading areas that will be targeted, the level of deficit students possess in the targeted areas, and the complexity of the areas to be targeted.

Current Practices in Tier 2 Interventions

We recently collected data on current practices in Tier 2 interventions in order to examine the extent to which the above research-based guidelines are being implemented for students with reading difficulties in the primary grades. First, the instructional components second-grade intervention teachers implemented in reading interventions for their students with reading difficulties were examined. Second, intervention delivery in terms of session length and frequency as well as instructional group size for general education interventions being implemented in kindergarten through Grade 3 were examined.

Content of Tier 2 Interventions

At Grade 2, Gersten et al. (2008), found research evidence for the implementation of phonics and vocabulary, and the panel additionally recommended fluency with connected text and comprehension as important targets for Tier 2 interventions. We examined the amount of time second-grade reading intervention teachers spent implementing each reading component during their Tier 2 sessions and the extent to which the implementation practices aligned with the research recommendations.

We collected data from all the reading intervention teachers for three elementary schools in three districts in three states. All of the schools were identified as Title I schools serving a high population of students at risk for reading difficulties. In addition, the schools had demonstrated instructional success meeting state reading standards, with a majority of students in all reporting categories meeting minimum passing standards on the state reading accountability measure the previous year. A total of five teachers were serving 20 second-grade students identified for Tier 2 intervention with oral reading fluency and/or comprehension screening measures. Four of the teachers exclusively provided Tier 2 reading intervention for various grade levels throughout the day, and one second-grade classroom teacher was responsible for providing Tier 2 intervention to several students in her class in addition to her Tier I instruction. Instructional group sizes for the Tier 2 interventions observed ranged from two to six students.

Data were collected on each teacher for three consecutive days in the fall semester of the school year. A three-day block was randomly assigned to each teacher across the semester with each of the 5 days of the week included with equal frequency over the entire data collection period. The Instructional Content Emphasis-Revised observation measure (ICE-R; Edmonds and Briggs 2003) was adapted and used to code the instructional elements of each intervention lesson.

We found teachers did focus their instruction on one to three reading components, usually phonics and word recognition, fluency, and/or comprehension. Phonics and word recognition was clearly the emphasis of instruction across most Tier 2 intervention groups with 66 % of observed time spent on sounds, decoding and encoding of words, and application to sentences. Only one of the instructional groups observed did not receive instruction in phonics and word recognition during the observations. Approximately 20 % of the total time observed was spent on fluency instruction and practice, with three of the observed lessons devoted exclusively to this component of reading. Specific time on comprehension instruction and practice was noted for approximately 7% of the observed time and typically involved students answering either oral or written questions about the text. No instruction in a specific comprehension skill or strategy was noted during this questioning; rather, the questions seemed to serve as comprehension checks or assessments of student understanding. Two observed lessons did include more specific comprehension instruction on story retell and the use of a graphic organizer to record story structure. Approximately 1% of the observed time included vocabulary instruction. This time amounted to a mean of 25 s of vocabulary instruction per lesson. Most of the vocabulary instruction that was observed occurred when a student asked about the meaning of a word during an activity, and did not appear to be a specific, targeted aspect of the lesson plan for the day. As previously reported, the majority of vocabulary instruction for students with reading difficulties seems to occur in Tier I instruction and is not often a targeted aspect of Tier 2 intervention in practice (Wanzek 2014).

In focus groups discussing their reading instruction, the intervention teachers emphasized the importance of phonics and word recognition and reading fluency work in their interventions, consistent with the observed instruction. Although several teachers noted the low vocabulary knowledge of the students they worked with, vocabulary as a targeted area of intervention was not mentioned by any of the intervention teachers. Overall, the observed intervention sessions mirrored most of the instructional recommendations of Gersten et al. (2008) for the content/emphasis of instruction for Tier 2 interventions, though we did not have student data to match student need to the content provided.

Delivery of Tier 2 Interventions

In terms of the delivery of Tier 2 interventions, current practice also seems to embody the recommendations of Gersten et al. (2008). In a recent study (Wanzek and Cavanaugh 2012), 1041 kindergarten through third-grade teachers regarding the general education interventions their students with reading difficulties received during the year were surveyed. At each grade level, the majority of teachers indicated students were scheduled in intervention daily. Only 18.7–29.5% of the teachers indicated they had students receiving intervention for 1 or 2 days a week. Intervention durations between 10 and 30 min were the most commonly reported. However, shorter interventions of 10–20 min were more common at the kindergarten level, and interventions of 31–40 min or 51 or more minutes were more commonly reported by third-grade teachers. Instructional group sizes of two to five students were the most commonly reported at each of the grade levels. The data on time, frequency, and instructional group size suggested that general education interventions were being implemented according to the research-based guidelines on the delivery of Tier 2 interventions provided by Gersten et al. (2008).

One finding that surprised us was the high number of Tier 2 interventions being provided by the classroom teacher. In fact, 73% of the classroom teachers indicated they were the provider of the supplemental intervention for at least some of their students with reading difficulties. This finding may be an effect of the current economic climate requiring many schools to work with fewer resources. A recent survey study of multi-tiered intervention in elementary schools across 17 states found approximately 37% of classroom teachers reported that they deliver Tier 2 intervention (Jenkins et al. in press) with 20% reporting Tier 2 occurred in the Tier 1 setting.

Current practice seems to be largely in line with current research regarding the instruction and delivery aspects of Tier 2 interventions at the primary grades. However, additional research is needed in relation to providing instruction in the components of vocabulary and comprehension as part of supplemental interventions. In addition, there is no research regarding the effects of Tier 2 interventions provided specifically by the general education classroom teachers, a practice that appears quite common currently in many schools throughout the primary grades. To provide an initial examination of this practice, the implementation of a hybrid Tier I and Tier 2 intervention provided by kindergarten classroom teachers and its effects for students identified as at risk for reading difficulties were investigated.

Examination of Hybrid Tier 1 and 2 Intervention

Al Otaiba and colleagues (2011) sought to implement an enhanced Tier 1 instruction that also provided the opportunity for classroom teachers to deliver Tier 2 intervention to students in their classes. In this study, kindergarten teachers in the treatment schools received professional development in differentiating instruction for students at risk for reading difficulties, providing additional time in instruction in small homogeneous groups, and targeting instruction to meet student needs. We examined the effects of this tiered intervention provided by classroom teachers for the students at risk for reading difficulties.

Study Implementation

Schools were recruited with assistance from the school district and were randomly assigned to the treatment (individualized instruction support) or comparison condition (typical practice). The 14 participating schools served an economically and ethnically diverse population of students. Schools were matched on criteria including proportion of students receiving free or reduced lunch, Title I and Reading First participation, and the state assigned school reading proficiency grades.

Teacher Participants A total of 44 kindergarten teachers participated in the study. The majority of the teachers were Caucasian ($n=31$, 70.5 %), 10 (22.7 %) were African American, and three (6.9 %) were Hispanic. All teachers in both conditions received a common baseline professional development consisting of a daylong researcher-delivered workshop on response to intervention, with a focus on effective Tier 1 instruction and differentiated instruction. Teachers at schools assigned to the treatment condition not only participated in the baseline professional development but also received supplemental training in implementing a data-informed hybrid of Tier 2 instruction and Tier 2 small-group, individualized interventions.

Student Participants As part of the larger study examining Tier 1 instruction for all learners, 605 kindergarten students provided parental consent. The students were screened in the fall of the school year using the Letter Naming Fluency (LNF) task of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good and Kaminski 2002). Students were selected for the current Tier 2 examination if they scored in the “high-risk” category (i.e., less than eight letters correct per minute) as defined by Good and Kaminski (2002). Alphabet knowledge and rapid naming, both measured on the LNF subtest, have been shown to have medium to large predictive relationships with later literacy development in the areas of decoding and comprehension (National Early Literacy Panel 2008). A total of 113 students (18.7 %) were identified as at risk for reading difficulties. Demographic data were not available for two of the students. Of the remaining sample, 58 (51.3 %) were male, and 53 (46.9 %) were female. The sample of kindergarten students was ethnically diverse, consisting of 74 African American (66.7 %), 28 Caucasian (25.2 %), 8 Hispanic (7.2 %), and 1 Asian (0.9 %). Eighty-nine students (78.8 %) were eligible for participation in the free or reduced-price lunch program. In addition, six (5.3 %) students were identified as limited English proficient (LEP). Approximately one third of the sample (35 students, 31.5 %) was eligible for special education services.

Sixty-seven students were in classrooms assigned to the treatment condition, and 46 students were in classrooms assigned to the comparison condition. There were no significant differences between the treatment and comparison conditions in gender, ethnicity, LEP, special education status, or age at time of testing ($p>.05$). However, the proportion of students participating in free or reduced lunch programs assigned to the treatment condition (87.7 %) was significantly higher than that of the contrast condition (69.6 %; $p = .018$).

Tier 2 Intervention Teachers in the treatment group were provided ongoing professional development, in-class support, and software (Assessment to Intervention [A2i]) for individualizing student instruction and providing small-group intervention to students at risk for reading difficulties in kindergarten. The A2i software (Connor et al. 2009; Connor et al. 2004) was designed to support teachers' ability to use individual student assessment data to inform decisions regarding instructional amounts, types, and groupings. Using child assessment data, specifically language and reading scores, the software applies fine-tuned algorithms that use a predetermined end-of-year target outcome to compute recommended content and amounts of instruction for individual students.

At-risk students received additional time in teacher-directed small-group instruction in recommended reading components based on the A2i software-based mathematical formulas. Using information derived from progress-monitoring information and student assessment data regarding individual students' strengths and weaknesses, teachers made specific decisions about the skills and strategies to target in the identified areas of Tier 2 intervention. Specifically, the A2i software suggested instructional approaches that were teacher- or child-managed and either code-focused (e.g., phonemic awareness, word recognition, fluency) or meaning-focused (e.g., vocabulary, comprehension). Decisions were also made by the teachers using this data concerning the amount of time per session, frequency of the sessions, group sizes, and planning for even further explicit, targeted instruction, practice, and feedback.

Throughout the school year, teachers received monthly professional development on individualizing instruction for students in their classes with a focus on student assessment data, scheduling the recommended minutes of intervention, and individualizing instruction in terms of the targeted content and delivery of instruction. Research staff followed up with each teacher with biweekly classroom-based support for intervention implementation. During biweekly visits, the research partners provided support to the teachers by reinforcing the professional development, assisting with technology as needed, modeling small-group strategies, and leading center activities.

Measures A set of early literacy and standardized reading achievement measures were administered to students throughout the school year. As stated previously, performance on the LNF task of the DIBELS was used to identify students as at risk for later reading difficulties at the beginning of kindergarten. This subtest is individually administered and consists of students being asked to identify as many randomly ordered upper and lowercase letters as they can in one minute. In addition, the Phoneme Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF) subtests of the DIBELS were administered in the winter and spring. PSF assesses student ability to orally identify individual phonemes in three- and four-phoneme words. NWF requires students to decode two- and three-letter pseudo-words, receiving one point for each letter sound produced correctly.

The Letter Word Identification, Word Attack, and Passage Comprehension subtests of the Woodcock Johnson III Achievement Tests (WJ-III; Woodcock et al. 2001) were administered to all students. The Letter Word Identification subtest was administered in the fall, winter, and spring, requiring students to identify letters and read words of increasing levels of difficulty. Word Attack was administered in the winter and spring. This subtest requires students to identify letter sounds and decode pseudo-words. Passage Comprehension was also administered in the fall, winter, and spring to assess students' ability to comprehend text from a short passage using the Passage Comprehension subtest of the WJ-III.

Findings

We examined whether the students at risk for reading difficulties who participated in a Tier

2 intervention provided by classroom teachers during their kindergarten year were able to increase their skills towards grade level expectations. The students in the treatment condition increased their accuracy and fluency in early literacy skills. At the end of kindergarten, students in the treatment condition had a mean of 38.13 correct letters per minute (SD = 23.10) compared to the comparison group mean of 30.57 correct letters per minute (SD = 16.49). These means represent an effect size in favor of the treatment of $d = 0.38$. However, the expected benchmark for end of kindergarten is 40 correct letters per minute. Students in the treatment condition reached a mean of 34.85 sounds per minute (SD = 28.66) by the end of the year on the PSF measure. In comparison, at-risk students not receiving the treatment scored a mean of 20.17 sounds per minute (SD = 16.04), representing an effect size of $d = 0.63$ in favor of the treatment. The expected benchmark for the end of kindergarten is 35 sounds per minute. On NWF, the at-risk students in the treatment group scored 27.54 letter sounds per minute (SD = 21.79). The at-risk students in the comparison group scored 18.11 letter sounds per minute (SD = 12.46). Consequently, the effect size for the treatment on NWF was $d = 0.53$. The expected benchmark at the end of kindergarten for NWF is 20 letter sounds per minute. By the end of the kindergarten year, mean scores for PSF and NWF were at or above expected levels for the treatment group while the comparison group fell well below expected levels on both measures. Overall, these data suggest at-risk students receiving the hybrid Tier 1 and 2 intervention provided by classroom teachers may be better prepared for literacy skills and print emphasis than the at-risk students whose teachers did not participate in the treatment.

Student scores on standardized measures of literacy skills (Letter Word Identification, Word Attack, and Comprehension) were also examined. These skills can be difficult to measure in kindergarten due to the limited reading levels of kindergarten students, particularly for students demonstrating risk for reading difficulties. However, these skills were measured in order to provide normative information regarding student trajectories towards important literacy skills. The W scores (Rasch ability scores with equal-interval measurement characteristics) and standard scores for each measure to allow for the best interpretative capabilities of student achievement with print and literacy were provided.

Small gains in letter word identification and comprehension were noted during kindergarten for all students as would be expected. At the end of kindergarten, at-risk students in both the treatment and comparison groups had standard scores in the average range on Letter Word Identification—mean scores of 95.61 (SD = 10.76) and 92.80 (SD = 12.58) for treatment and comparison, respectively—despite beginning the year with below average skills (treatment $M = 84.10$ [SD = 8.78]; comparison $M = 82.15$ [SD = 9.61]). As a point of comparison, these beginning-of-the-year standard scores represent *W* scores of 324.43 (SD = 16.70) and 322.28 (SD = 15.80) for the treatment and comparison groups, respectively. The end-of-the-year *W* scores were 371.82 (SD = 19.26) and 367.54 (SD = 20.15) for the treatment and comparison, respectively. Taking pretest scores into account, a small effect size of $d = 0.12$ was noted in favor of the treatment on Letter Word Identification at the end of the year.

End-of-year Word Attack scores showed a similar pattern with the treatment group achieving an average standard score of 100.57 (standard deviation, SD = 12.61), or an average *W* score of 425.07 (SD = 24.96); the comparison group achieved an average standard score of 96.17 (SD = 14.26), equivalent to an average *W* score of 417.22 (SD = 26.22).

By the end of kindergarten mean standard scores for students in both the treatment and comparison groups in comprehension were in the average range with mean scores of 91.55 (SD = 10.94) and 88.41 (SD = 11.89) for the treatment and comparison, respectively. These standard scores represent *W* scores of 410.42 (SD = 14.53) and 406.91 (SD = 16.13) for the treatment and comparison, respectively. Taking pretest into account a small effect size of $d = 0.18$ was noted in favor of the treatment on Passage Comprehension.

Fig. 1 Letter Naming Fluency for kindergarten students

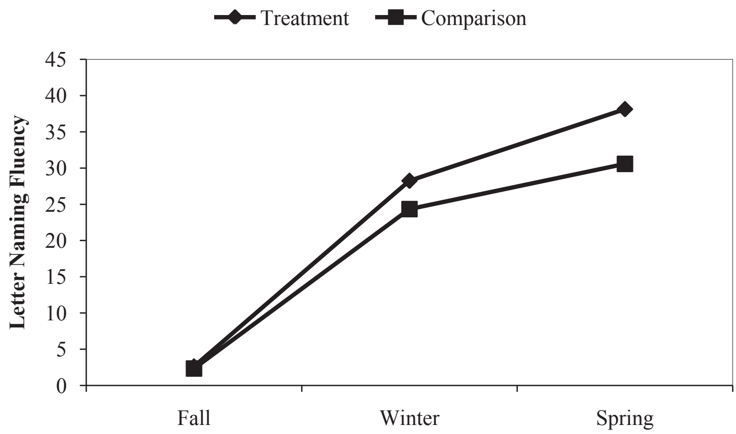


Fig. 2 Phoneme Segmentation Fluency for kindergarten students

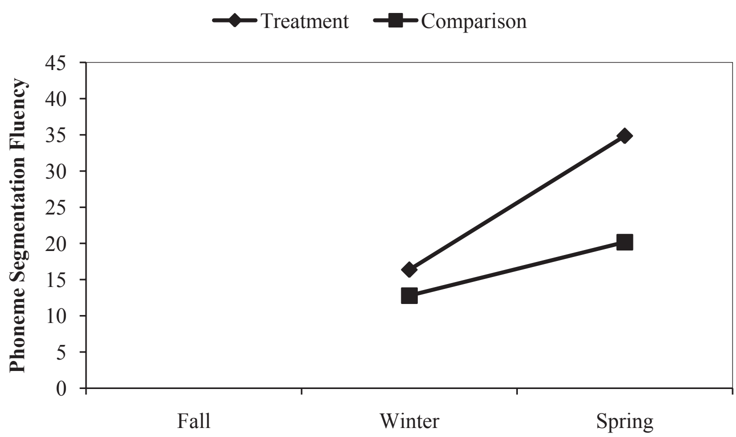
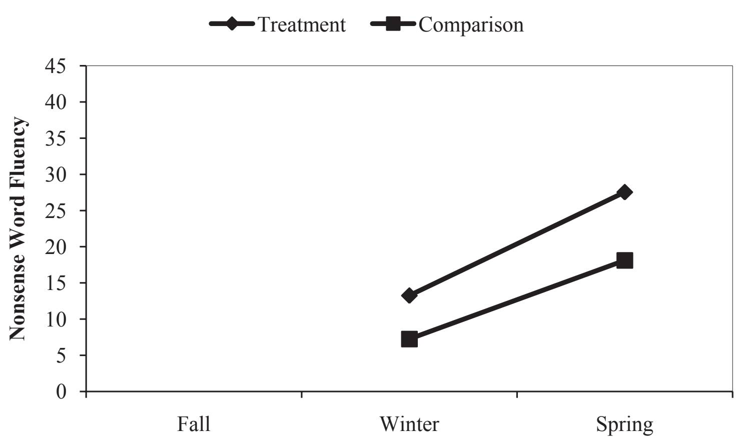


Fig. 3 Nonsense Word Fluency for kindergarten students



Figs. 1, 2, 3, 4, 5, and 6 display the relative performance over time on each measure for the treatment and comparison groups. The graphs project a consistent pattern of higher student response to intervention for students participating in the Tier 2 intervention. However, as noted in the effect sizes above, smaller effects were noted on the standardized measures of literacy

Fig. 4 Letter Word Identification mean *W* scores for kindergarten students

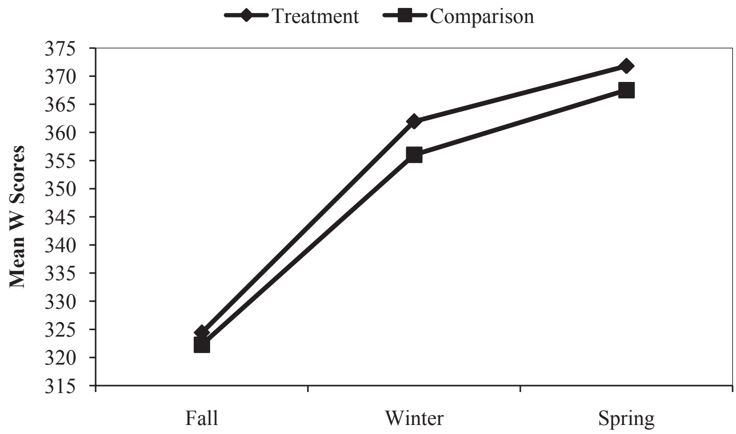


Fig. 5 Word Attack mean *W* scores for kindergarten students

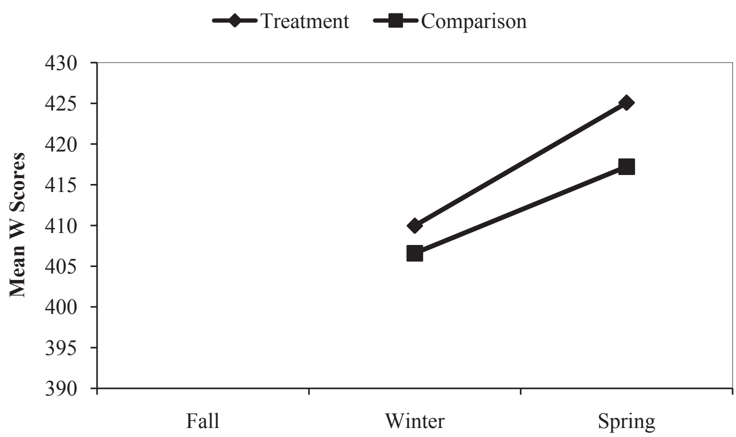
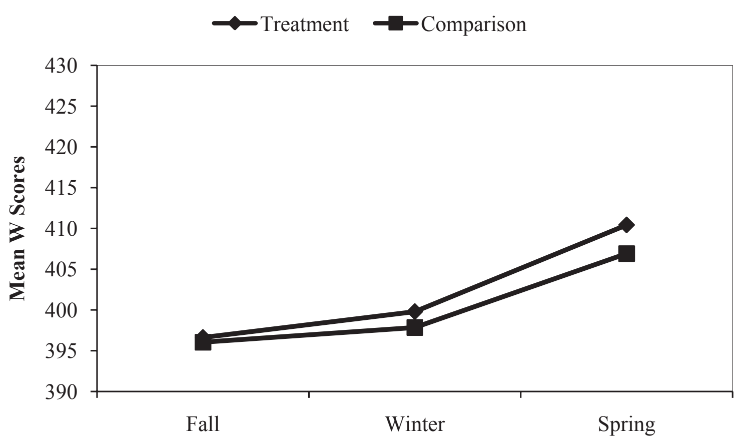


Fig. 6 Passage Comprehension mean *W* scores for kindergarten students



skills than on the early literacy skills. In addition, the standard scores were in the average range for each measure, though comprehension standard scores were lower than the word reading scores. Thus, although these students' skills may not yet be strong enough to ensure Tier I instruction alone, without continued Tier 2 support, they can sustain their reading growth in the coming grade levels.

Table 1 Summary of Tier 2 recommendations and current practice

Research-based recommendations for Tier 2 implementation	Current practices for Tier 2 implementation
Intensive, systematic instruction on up to three foundational reading skills	Instruction is focused on 1–3 reading components At Grade 2, phonics and word recognition, fluency, and/or comprehension were noted Many classroom teachers are implementing instruction for Tier 2 intervention
Provide instruction in small groups	Most interventions implemented in groups of 2–5 students
Typically, these groups meet 3–5 times per week for 20–40 min	Most interventions are 4–5 days per week Most interventions are 10–30 min in length Shorter interventions are noted in earlier grades (kindergarten) with longer interventions noted at third grade

Implications and Future Research

Implementation of Tier 2 interventions is buoyed by a solid research base, particularly at the early elementary level. Providing targeted instruction in one or more early reading skills in small instructional groups for 20–40-min sessions to match student needs can assist many students in improving their reading achievement and meeting grade level expectations. It is encouraging that many of the research-recommended components of effective Tier 2 interventions are being implemented in practice. Table 1 provides current research-based recommendations for implementing Tier 2 interventions in a response to intervention model.

Currently, in practice, vocabulary and/or comprehension instruction do not seem to be an emphasis of Tier 2 interventions implemented at the early elementary level. In fact, the role of vocabulary and comprehension instruction in Tier 2 interventions has received the least attention in the literature for the early grade levels and some have suggested students with reading difficulties in the early grades should receive this instruction primarily through Tier 1 (Gersten et al. 2008). Vocabulary and comprehension become increasingly emphasized through the grade levels and early difficulties in these areas predict future reading problems (Catts et al. 2006; National Early Literacy Panel 2008). Thus, additional research on effective instruction for students at risk for reading difficulties including information on the emphasis of the different components of reading matched with student need in the various tiers of a response to intervention model is needed.

Currently, many classroom teachers are charged with implementing Tier 2 interventions for their students. Assuming that classroom teachers can effectively implement individualized small-group instruction is problematic. Spear-Swerling and Cheesman (2011) examined elementary school teachers' knowledge about evidence-based reading instruction, assessment, and response to intervention. Teachers showed limited knowledge about assessment and about interventions suitable for Tier 2 within a response to intervention framework, although teachers with more professional development had significantly higher performance than teachers without such training. To date, the research base on Tier 2 interventions is built on supplemental interventions provided by other personnel or researchers without the scheduling and instructional demands of managing interventions for struggling readers within a full general education class of learners. However, with current school resources, Tier 2 interventions implemented by classroom teachers may be the only option in some cases, allowing more specialized personnel to focus on implementation of Tier 3 interventions. Our research in kindergarten suggests students can improve in their reading achievement with these classroom-teacher implemented Tier 2 interventions when their teachers are provided training and support. However, there was no direct comparison to Tier 2 interventions provided outside of the classroom or by other personnel so whether students would have improved even further in these types on Tier 2 interventions cannot be said. Certainly, the students who showed slow response to the intervention may have benefitted from a more intense intervention that was provided either outside the classroom, in smaller groups, or for additional time. Furthermore, given the relatively weaker comprehension outcomes, it is possible that students would have benefitted from more explicit vocabulary and comprehension instruction.

A frequent difficulty with classroom teachers implementing the Tier 2 intervention is the amount of time needed to implement the intervention and also provide meaningful activities for students in the class who are not participating in the intervention. Future research examining the practical conditions under which Tier 2 interventions can be effectively implemented could help schools with varying resources make important, cost-effective decisions regarding these interventions. Related, there is still limited research on the timing of placement in Tier 2 versus Tier 3 interventions. Although it is expected at the early elementary grades that students with reading difficulties participate in Tier 2 interventions and receive Tier 3 interventions only after demonstrating insufficient response to Tier 2, the point at which a student's reading growth should be considered insufficient remains imprecise. In addition, the extent to which there are students, even in the earliest grades, who would benefit most from immediate placement in Tier 3 interventions has not yet been studied.

Funding Information

This research was supported by Grant 1R03HD060758-01A1 and Grant P50HD052120 from the National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Child Health and Human Development or the National Institutes of Health.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal, 111,* 535–560.

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S., Shaywitz, B., & Shaywitz, S. (2004). Effects of intensive reading remediation for second and third graders. *Journal of Educational Psychology, 96,* 444–461.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49,* 278–293.

Connor, C. M., Morrison, F. J., & Katch, E. L. (2004). Beyond the reading wars: The effect of classroom instruction by child interaction on early reading. *Scientific Studies of Reading, 8,* 305–336.

Connor, C. M., Piasta, S. B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P., & Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions of first graders' literacy development. *Child Development, 80,* 77–100.

Denton, C. A., Fletcher, J. M., Anthony, J. L., & Francis, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities, 35,* 447–466.

Edmonds, M., & Briggs, K. L. (2003). The instructional content emphasis instrument: Observations of reading instruction. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observation of teaching and learning* (pp. 31–52). Baltimore: Brookes.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene: Institute for Development of Educational Achievement.

Jenkins, J. R., Schiller, E., Blackorby, J., Kalb Thayer, S., & Tilly, W. D. (in press). Response to intervention in reading: Architecture and practices. *Learning Disabilities Quarterly*.

Luftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice, 25,* 124–136.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71,* 445–463.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the national early literacy panel*. Washington, DC: National Center for Family Literacy.

O'Conner, R. E., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and student outcomes. *Journal of Learning Disabilities, 38,* 440–455.

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research and Practice, 25,* 110–123.

Spear-Swerling, L., & Cheeseman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing: An Interdisciplinary Journal, 25,* 1691–1723.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593.

Tuckwiller, E. D., Pullen, P. C., & Coyne, M. D. (2010). The use of the regression discontinuity design in tiered intervention research: A pilot study exploring vocabulary instruction for at-risk kindergarteners. *Learning Disabilities Research & Practice, 25,* 137–150.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75,* 165–183.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early interventions as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–38.

Wanzek, J. (2014). Building word knowledge: Opportunities for direct vocabulary instruction in general education for students with reading difficulties. *Reading and Writing Quarterly, 30,* 139–164.

Wanzek, J., & Cavanaugh, C. (2012). Characteristics of general education reading interventions implemented in elementary schools for students with reading difficulties. *Remedial and Special Education, 33,* 192–202.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca: Riverside Publishing.

Part V

Tier 3– Assessment, Problem Analysis, and Intervention

Progress Monitoring for Students Receiving Intensive Academic Intervention

David A. Klingbeil, Tera L. Bradley and Jennifer J. McComas

Progress monitoring, the repeated measurement of academic skills to inform instruction (Shapiro 2008), is one of the four essential components for response-to-intervention (RTI) systems (National Center on Response to Intervention, (NCRTI) 2010). Assessment data can be used to measure progress toward individual goals or state standards, monitor intervention effectiveness, identify instructional objectives, and identify students who need additional support (Salvia et al. 2013). Monitoring student progress and subsequently modifying instruction can lead to increased academic achievement for struggling learners (Fuchs et al. 1984; Fuchs et al. 1989; Fuchs et al. 1991; Stecker and Fuchs 2000). Research conducted over recent years has significant implications for collecting and interpreting progress-monitoring data for students receiving intensive academic support.

In this chapter, individual progress monitoring for students needing intensive interventions in academics is discussed. First, the general purposes of progress monitoring students receiving tier 3 interventions and highlight differences from progress monitoring within other tiers are described. Second, two types of data that are useful for academic progress monitoring in an RTI model are reviewed. Third, the practical and technical considerations necessary for progress monitoring within tier 3 are discussed. Fourth, case examples to illustrate the types of decisions made for students receiving tier 3 interventions are provided. Finally, the chapter concludes with a summary of recommendations for practice and a discussion on directions for future research.

D. A. Klingbeil (✉)
University of Wisconsin-Milwaukee, Milwaukee, USA
e-mail: davidak5@uwm.edu

T. L. Bradley
Boling Center for Developmental Disabilities, Memphis, USA

J. J. McComas
University of Minnesota, Minneapolis, USA

Purpose of Progress Monitoring

Student progress is measured repeatedly within RTI systems to assess students' academic performance, measure skill growth over time, and evaluate intervention effectiveness (NCRTI 2010; Shapiro 2008). Progress-monitoring data must be collected at least monthly (Fuchs and Fuchs 2011; NCRTI 2010; Stecker et al. 2008) and can be used to inform a variety of decisions across RTI tiers. At tier 1, progress-monitoring data can be used to monitor the effectiveness of class-wide instruction. Combining progress-monitoring procedures with universal screening data leads to more accurate identification of students needing additional intervention (Compton et al. 2010; Compton et al. 2006).

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_20

The purposes of progress monitoring, and subsequent requirements for assessment procedures, change as intervention intensity increases across tiers. To begin, all students' academic progress is measured approximately three times per year with a benchmark assessment. Students who are performing slightly below grade level may require a short-term, moderately intense evidence-based intervention (i.e., tier 2; NCRTI 2010). The frequency of progress monitoring increases at tier 2, with most researchers recommending weekly data collection (Burns and Gibbons 2012; Hintze and Marcotte 2010). Educators use these data to inform decisions about intervention effectiveness and whether the student requires more intensive intervention.

Tier 3

There are a few unique considerations for monitoring the progress of students receiving intensive (i.e., tier 3) interventions. Students receiving intensive interventions have persistent and severe learning problems in one or more academic skill areas (National Center on Intensive Intervention (NCII) 2013). Data to evaluate tier 2 interventions are usually collected with grade-level material, but student progress may be measured with below grade-level (i.e., instructional level) materials in tier 3, depending on the magnitude of the academic skill problems (Shapiro 2011). Diagnostic assessments, error analysis, or brief experimental analyses may also be used to pinpoint specific skills needing remediation and identify potential intervention. Given the further intensification of the intervention, student progress is monitored more frequently to formatively evaluate instructional effectiveness (NCII 2013; Stecker et al. 2008).

Decisions made based on progress-monitoring data can also differ within tier 3. If students are not making progress, the intervention should be modified to meet individual learning needs of the student. Theoretically, support at this level is the most intensive possible in regular education. If the further intensification and modification is not successful in remediating the skill deficits, educators may use these data to initiate a special education evaluation. Progress-monitoring data may also be used as part of special education eligibility determinations (Vaughn and Fuchs 2003, 2012). The severity of academic problems and the high-stakes nature of the decisions made for students receiving intensive interventions require the most exacting assessment practices. Such high-stakes decisions also necessitate strict scrutiny of the evidence for the procedures, measures, and interpretations of the resulting data.

Data to Inform Tier 3 Decision-Making

Suitable methods for progress monitoring within an RTI system provide direct observations of the targeted skill, are sensitive to small changes in skill, and result in data that inform instructional decisions (Stecker et al. 2008; Ysseldyke et al. 2010). Assessment procedures must be time efficient and easy to administer reliably. Selected measures must also have an adequate number of alternate forms and evidence of psychometric adequacy for the proposed uses of the data (Riley-Tillman et al. 2013; Stecker et al. 2008). Educators use a variety of instruments for progress monitoring (Mellard et al. 2009). Curriculum-based assessment methods are widely described as appropriate for use for individual progress monitoring.

Curriculum-based assessment (CBA) was designed to collect instructionally relevant assessment information by frequently measuring student performance in the classroom curriculum (Tucker 1985). Researchers created guidelines for educators to create their own materials in order to maximize the alignment with classroom instruction (Shinn 1989; Fuchs and Deno 1991). There are multiple methods of CBA but these approaches generally incorporate one of two models of assessment: mastery measurement and general outcome measurement (Fuchs and Deno 1991).

Mastery Measurement Many approaches to conducting CBA emphasize subskill mastery measurement (Hintze et al. 2006). Mastery data are useful for planning instruction and measuring progress toward short-term intervention objectives (Shapiro 2011). Subskills are identified by separating broad curricular objectives, or skill areas, into a hierarchy of discrete skills (Fuchs and Deno 1991). For example, the broad area of reading can be broken down into the core areas of phonemic awareness, phonics, fluency, vocabulary, and comprehension (National Reading Panel 2000). Phonics skills can be separated further into a student's ability to decode consonant–vowel–consonant patterns, diphthongs or diagraphs, long vowel sounds, etc. Mastery measures align with current instructional targets to assess students' skills in a narrowly defined area. Educators often create their own mastery measures but other options exist.

Mastery measures should be brief and easy to administer. It is important that there are enough assessment items for the student to demonstrate mastery but students will likely reach a ceiling as skills are mastered (Christ and Vining 2006; Hosp and Ardoin 2008). Items within one mastery measure or "probe" should be of relatively equal difficulty. Resulting data provide a measure of student learning for a specific task. Mastery measures should provide information about the student's accuracy and fluency with the skill as both are required for skill proficiency (Hosp et al. 2014). When the instructional targets change, a new probe is required to measure the skill. Inferences regarding student performance across mastery measures are not permissible as the difficulty of mastery probes shift based on task complexity (Fuchs and Deno 1991; Hintze et al. 2006).

General Outcome Measurement General outcome measures (GOM) provide a standardized method for determining student progress toward long-term goals such as end-of-year learning targets or state standards (Shapiro 2011). This requires brief measures that contain a random sample of all the skills taught in a curriculum. Including a sample of all the skills in a curriculum also provides an index of retention and generalization (Fuchs and Deno 1991). This information is not captured by mastery measures. Data from general outcome measures are generally predictive of performance on standardized achievement tests and statewide assessments (Keller-Margulis et al. 2008). These measures are less sensitive to instructional effects over a very brief period by design.

Curriculum-based measurement (CBM) is a popular and well-researched type of GOM. CBM was developed to help special educators formatively evaluate instructional effectiveness (Deno 1985). The process involved systematic sampling of the yearly curriculum to create brief assessments. Today, most GOM assessments are developed to measure robust indicators of academic skill (Fuchs 2004) and may be aligned with state standards (e.g., University of Oregon Center for Teaching and Learning 2012). These curriculum-independent measures are still often described as CBM, which is a likely source of confusion. The importance of selecting assessment materials is discussed below, but the term GOM can be used to describe assessments that incorporate similar measurement procedures as CBM (Fuchs and Deno 1994).

Collecting GOM data requires the use of standardized data collection and scoring procedures. Students are assessed with brief tests or "probes" that are of equal difficulty. Ensuring the equal difficulty across all of the probes is essential for assessing learning in a short period of time. Measuring progress over time also requires a sufficient number of alternate forms. These data are then graphed in a time-series fashion to provide an indication of students' progress toward long-term goals.

The purposes for underlying mastery measures and GOM data differ, but educators can collect these data using similar procedures. Procedures associated with general outcome measures (e.g., using timed assessments and standardized scoring) can be used when collecting information about skill mastery (Hosp et al. 2014). Indeed, CBM tools and procedures have been used to collect mastery measures across several academic skills (Hintze et al. 2002; Hosp and Fuchs 2005; Shinn 1989). For example, educators could use single-skill math CBM probes to collect mastery data and multi-skill math CBM probes to collect

GOM data (Burns and Klingbeil 2010; Christ and Vining 2006). Both types of data are necessary for monitoring the progress of students receiving intensive supports (Shapiro 2011).

Evidence for Monitoring Progress within RTI Systems

Educators must make multiple decisions when monitoring student progress including: (a) the selection of progress-monitoring tools, (b) establishing procedures to measure baseline performance, (c) duration and schedule of data collection, or (d) rules for interpreting progress-monitoring data. Readers familiar with literature in this area have likely read multiple rules and best-practice recommendations for making these decisions. Recommendations for practice should maximize the psychometric adequacy of assessment procedures (Salvia et al. 2013), which will improve the accuracy of the data-based decisions made by educators. Establishing psychometrically defensible practices is an on-going process (Messick 1989). Educators must take care to implement practices that are based on the most current research.

General outcome measures, and in particular CBM-reading (CBM-R) scores, are widely described as psychometrically adequate and empirically valid for monitoring progress (e.g., Deno 2003; Hosp et al. 2008). Evidence of the psychometric adequacy of CBM scores is usually presented in terms of traditional definitions of reliability and validity. There is a robust body of evidence documenting the test-retest (i.e., measurement consistency across two time points), alternate form (i.e., measurement consistency across forms of an assessment), and inter-rater (i.e., reliability across administrators) reliability of CBM data (Wayman et al. 2007).

Discussions regarding the validity of CBM measures focus on concurrent and predictive relationships between CBM data and other measures. The relationships between CBM and standardized tests of achievement have been well established in reading (e.g., Marston 1989; Reschly et al 2009), writing (McMaster and Espin 2007), and math (Foegen et al. 2007). Validity evidence is also drawn from correlations between CBM data and criterion-referenced state assessments (e.g., Hintze and Silberglitt 2005; Shapiro et al. 2006; Stage and Jacobsen 2001). Evidence suggests that the collection of progress-monitoring data to inform instructional modifications results in improved achievement for students requiring intensive support (Fuchs and Fuchs 1986; Fuchs et al. 1991). Taken together, these findings are widely interpreted to suggest that GOM data can be used to monitor student progress across RTI tiers.

More recent conceptualizations of validity focus on the proposed uses and interpretations of assessment data (Messick 1995). According to Kane (2013a, p. 1) "the validity of a proposed interpretation or use of test scores can be defined in terms of the plausibility and appropriateness of the proposed use." Kane (2013a) identified four points that educators should consider when reviewing the psychometric evidence for any assessment practice. First, proposed interpretations and uses of test scores are validated, not the test. Second, scores from the same test can be valid for some purposes and invalid for others. Third, validity may change over time as new interpretations and uses are proposed. Fourth, evidence of reliability is subsumed within the validation process but reliability is not sufficient for validating score interpretations. Points 2 and 3 are especially relevant for the focus of this chapter.

Low-stakes decisions (e.g., instructional modification) generally require lower evidence for psychometric adequacy (e.g., reliability of .70), whereas high-stakes decisions (e.g., retention, entitlement eligibility) require that higher psychometric standards be met (e.g., reliability of 90; Salvia et al. 2013). These estimates must be related to the proposed uses of the data. The evidence of reliability and validity of GOM data reviewed above is not enough to ascertain that the measures adequately capture growth (Christ et al. 2012; Deno et al. 2001; Francis et al. 2008).

Effective progress monitoring requires assessment data that capture student progress toward short-term and long-term learning goals (Shapiro 2011). The steps for monitoring progress are fairly straightforward. Educators must establish baseline performance, set goals for improvement in response to intervention, and analyze the data to make decisions (NCII 2013). Practical requirements for monitoring progress (e.g., passage selection, length or schedule of data collection) depend on decisions being made from these data.

Has the Student Learned What Was Taught?

Using progress-monitoring data to assess immediate skill acquisition is consistent with a mastery-measure approach to assessment (Fuchs and Deno 1991; Shapiro 2011). Decisions regarding skill acquisition pertain to criteria for determining student proficiency with a task. Mastery measures assess a narrow domain of behavior and students often quickly reach a ceiling as the skill is acquired (Christ and Vining 2006).

Previous Research Mastery should be determined by comparison to criteria that are "empirically determined and validated" (Hosp and Ardoin 2008, p. 74). Criteria for determining instructional level appear useful for this task, and are used to determine the proper difficulty level of academic tasks for individual students (Gickling and Thompson 1985). Tasks that are too hard may frustrate the learner while tasks that are too easy may not provide enough challenge. Teaching at the instructional level has been associated with increased achievement and time on task (e.g., Gickling and Armstrong 1978; Treptow et al. 2007). Instructional levels are best defined using accuracy (e.g., 93–97 % words read correctly for reading and 90 % known for other academic tasks; Burns, 2004) and fluency criteria (e.g., words read correct per minute; Burns et al. 2006).

Students should be taught specific skills using instructional level materials until their performance reaches the mastery level (Riley-Tillman et al. 2013). Consider a fourth-grade student who is receiving intensive intervention targeting two-digit subtraction with regrouping. Using the criteria derived by Burns et al. (2006), this skill could be taught until the student scores higher than 49—digits correct per minute (DCPM) with 90 % accuracy on a measure targeting only that skill. At that point, instruction would switch to another skill-requiring remediation. Empirically derived instructional level criteria for reading and early writing skills (Parker et al. 2011) could be used similarly. For reading, accuracy criteria of correctly reading 93–97 % of the words in context have a strong research base (Burns 2007; Gickling and Armstrong 1978; Treptow et al. 2007).

Mastery measures should be evaluated by their psychometric adequacy, sensitivity to student improvement, rules for determining mastery, and decision rules for changing instruction (NCII 2013). Although there has not been as much empirical research of published or locally created mastery measures as CBM, the research that has been conducted has found psychometrically adequate data (e.g., Burns 2001; Burns et al. 2000; Burns and Mosack 2005). Moreover, empirical and theoretical support for formatively evaluating student learning does underscore the importance of collecting mastery data to inform instructional modifications. The narrow focus of mastery measures does not permit inferences to broader skill domains (i.e., extrapolation inferences, Kane 2013b) or inferences regarding a student's ability to respond to an evidence-based intervention. These data have a limited association with overall achievement and should not be the sole criterion for making decisions regarding the continuation or discontinuation of intensive intervention.

Implications for Practice Intensive interventions are highly individualized to target students' specific skill deficits. Mastery measures are typically used to assess student progress toward a narrowly defined instructional target. Educators have the option to create their own mastery measures or use one of the few commercially developed measures. For example, single-skill math CBM probes that align with instructional targets could be used to measure skill mastery (Burns and Klingbeil 2010). Similarly, measures of early reading skills such as nonsense word fluency or letter sound fluency could be used to collect mastery data for older students performing below grade level.

Educators should provide instruction on the skill until data suggest the student is proficient (i.e., fluent and accurate). Previous research identified 93–97% accuracy as a criterion for measures that include reading connected text (Gickling and Thompson 1985) and 90% could be an accuracy criterion for most other skills (e.g., math fact, spelling, letter sounds; Burns 2004), but fluency criteria are less clear. An example mastery-level criterion could be the year-end benchmark for the grade-level that represents the skill being taught (Shapiro 2011). For example, a fourth-grade student receiving a decoding intervention could have his progress monitored with nonsense word fluency, and the beginning second-grade benchmark score could be used to determine proficiency because that is the highest grade level at which nonsense word fluency is used as an indicator of proficiency (University of Oregon Center on Teaching and Learning 2012). More information on the use of instructional levels to denote mastery is provided by Shapiro (2011) and Burns and Parker (2014).

Is the Student Making Progress Toward Grade-Level Objectives?

Mastery measures are not suitable for assessing growth over time because student performance does not translate across individual skills. General outcome measures are better suited to estimate growth because they sample behavior in a broad skill area (or across the entire curriculum). Measuring growth in this area requires multiple assessments of equal difficulty that are delivered throughout the year (Fuchs and Deno 1991). General outcome measures, in particular CBM, are widely used to monitor student progress in RTI systems. This requires reliable data that are sensitive to student growth over a short period of time.

Previous Research Most evidence for the sensitivity of GOM data, in particular CBM-R, comes from studies targeting groups of students. For example, Deno et al. (2001) used regression methods to create growth estimates based on grades and special education status. Hintze et al. (1994) found that CBM-R data were sensitive to differences in growth for groups of students exposed to different curricula. These studies do not provide evidence that these measures are sensitive to growth for an individual student (Ardoin et al. 2013). Unfortunately, more recent research has found instability in individual-level growth estimation (e.g., Ardoin and Christ 2009;Christ et al. 2013c, Francis et al. 2008; Hintze and Christ 2004).

Research in this area often uses complex procedures and terms. Generally, research on the estimation of student progress focuses on (a) observed performance, (b) true performance, (c) observed rates of growth, and (d) true rates of growth (Francis et al. 2008). Observed performance is captured by a single observation whereas observed growth is calculated using ordinary least squares slopes. True performance and true growth are latent constructs that are never known. Measure sensitivity quantifies the difference between observed rates and true rates of growth. Researchers have systematically evaluated the influence of baseline estimation; duration and schedule of data collection; and data quality on the reliability and sensitivity of growth estimates. Research on these factors is briefly summarized and implications for practice are discussed in the next sections. But first, a few considerations for the reader to keep in mind are discussed.

Considerations and Implications of Progress-Monitoring Research

Much of the research described below was conducted via computer simulation methods (Christ et al. 2012, 2013c; Van Norman et al. 2013) with some notable exceptions (e.g., Ardoin and Christ 2009; Hintze and Christ 2004; Thornblad and Christ 2014; Yeo et al. 2012). Computer simulation methods use existing data parameters to create an immense database to test hundreds of conditions that would not be feasible given the time and monetary requirements in schools. For example, the series of studies by Christ and colleagues used data from approximately 3100 second- and third-grade students to simulate a database with 9000 students per condition. Simulation methods are limited, as readily stated by the researchers, due to the estimation of fixed and random effects based on the field-based data. To minimize this threat, Christ et al. (2012, 2013c) compared the parameter estimates with those from field-based studies. However, replications of these findings derived from other field-based data sets are necessary. These findings are also specific to CBM-R scores and should not be generalized to other GOM assessments. Readers are urged to keep these limitations in mind while interpreting the findings.

Baseline Performance

Establishing a valid baseline level of performance is an important aspect of progress monitoring. Baseline data should be collected with a GOM that is aligned with the long-term intervention goals. These data are used to set a goal for skill improvement, which is essential for making decisions based on student performance at a later time.

Previous Research Using GOM data to establish baseline performance requires the inference that student scores accurately capture current skill levels. Unstable or inaccurate baseline estimates may bias the evaluation of student growth in response to intensive intervention (Jenkins et al. 2009). For example, underestimating student performance at baseline may lead to the overestimate of intervention effects. This could delay the modification of a marginally effective intervention. Yet, extending the baseline period will delay intervention implementation. Educators must balance the need for accurate estimations with the need to intervene quickly.

Recent studies provide guidance for using CBM-R data to estimate baseline performance. Collecting only one observation (e.g., student is assessed with one CBM-R passage for 1 min) may bias the precision of growth estimates, especially when data are collected for fewer than 8 weeks (Jenkins et al. 2009; Van Norman et al. 2013). When progress monitoring is conducted over a longer period, baseline estimation has a smaller impact on the accuracy of progress monitoring. Results from these studies suggest that educators can sufficiently estimate baseline performance for calculating student growth using the median of three observations (e.g., student is assessed with three CBM-R passages) collected at one time. Each of these studies focused on the accurate estimation of oral reading fluency. Additional research is needed to determine the extent to which these results hold for GOM measures in other skill areas.

Implications for Practice Educators should estimate baseline performance using a GOM that represents long-term instructional targets. Baseline estimation with GOM data requires using the median score from three assessments delivered at one time. Use of a median score should guard against bias from potential under- or over-estimation from a single observation and does not seem to be overly time-intensive.

Data Quality

According to Christ et al. (2012), data quality is related to the technical characteristics of the measure and the procedures used to collect the data. Original CBM procedures required random sampling of skills from the student's curriculum (Fuchs and Deno 1991). An important weakness in this approach was the variable difficulty of these locally created measures (Fuchs and Deno 1994). Researchers have developed standardized GOM assessments to control difficulty across the alternate forms required for measuring progress. The nontrivial error that remains is an important consideration.

Previous Research Multiple studies have found that student performance can vary widely across CBM-R probes. Estimates of variability range from 11 to 26 words read correct per minute (WRC/M), even when passages were controlled for difficulty (Ardoin and Christ 2009; Francis et al. 2008; Hintze and Christ 2004; Poncy et al.

2005). Variability in student performance appears to differ across well-developed and frequently used passage sets as well. A recent field-based study found that the error in point estimates ranged from 11 to 15 WRC/M depending on the passage set used (Ardoin and Christ 2009). Consider a student who read 100 words correctly in 1 min. We would be 68% confident that a student's true WRC/M ranged from approximately 89 to 111, when the standard error of the estimate was 11. Preliminary evidence suggests that the reliability of each data point differs over time as well (Yeo et al. 2012).

Measure selection also influences the estimation of student progress over time. Students can be expected to progress approximately 1–2 WRC per week (Deno et al. 2001). Research suggests that the error in slope estimates may approach the expected value of growth (Christ 2006; Hintze and Christ 2004). As with point estimation, errors in slope estimates differ across passage sets (Ardoin and Christ 2009).

Data collection procedures may also impact the quality of progress-monitoring data (Christ et al. 2012). Most general outcome measures have standardized instructions and scoring procedures, but it is unclear how much these vary in practice. Evidence is somewhat divergent on the effect of departures from standardization on student performance (Christ et al. 2013c; Colón and Kranzler 2006; Taylor et al. 2013). More research is needed to further estimate the impact of data collection on student performance and progress-monitoring data quality, especially with students needing intensive intervention.

Implications for Practice An important implication from this research is that measures used may account for more variability in student performance than the student's true growth (Ardoin and Christ 2009). Generating materials from the curriculum has face validity but will likely result in data with a large amount of error (Hintze and Christ 2004). Error may be lessened by using well-researched assessment tools, but it remains a critical consideration. Departures from standardization procedures may also reduce the quality of progress-monitoring data. Initial estimates for data quality suggest that error estimates of 5 WRC/M for CBM-R represent very good quality data, error estimates of 10 would be of good quality, and error estimates of 15 would be of poor quality (Christ et al. 2012). Obtaining high-quality data may be difficult in applied settings.

Research regarding the influence of measure quality on growth estimation has focused on CBM-R data. These results do not generalize to other GOM measures or CBM in other skill areas. The results do present reason for caution when interpreting the error that may be present in any progress-monitoring measure. Standard error of the estimate and standard errors of the slope values should be central considerations when selecting progress-monitoring tools. The NCII publishes annual ratings of common commercially developed GOM across a variety of areas which are available online at no cost (http://www.intensiveintervention.org). Educators can easily access up-to-date information regarding the utility of various progress-monitoring tools.

Duration and Schedule

Questions regarding the amount of data needed and the length of progress monitoring are of critical importance to educators. Ardoin et al. (2013) reviewed approximately 80 manuals and empirical articles that discussed the number of data points necessary to make a decision regarding student progress. The most common recommendation was seven data points, followed by six or ten data points. Moreover, these recommendations had limited empirical support. Recent evaluation of individual growth estimates suggests that inferences based on this number of data points would likely be inaccurate (Christ 2006).

Previous Research Recommendations regarding the number of data points to estimate growth reflect the duration (i.e., number of weeks) and the schedule (assessments per occasion, number per week) of progress monitoring. Christ (2006) recommended that progress monitoring occur for 10 weeks with data collected twice per week based on the parameters of published field-based studies. Jenkins et al. (2009) examined the impact of reducing the schedule of progress monitoring across a fixed duration of 10 weeks. Results suggested that the data collection could be thinned as long as multiple observations were collected. Notably, they found that collecting data only once per week raised the slope estimates. The reliability of these growth estimates did not reach psychometric standards for making high-stakes decisions, regardless of the schedule of data collection. Jenkins et al. (2009) did not examine the influence of data quality.

Christ et al. (2012, 2013c) used simulation methods to examine the influence of duration, the schedule of data collection, and the quality of the data set. The findings of Christ et al. 2013c) with a specific focus on results based on good quality (residual error = 10) or poor quality (residual error = 15) data are briefly summarized. These data sets may be more common in practice than very good-quality data (residual error = 5). It bears repeating that these findings can only be generalized to the use of CBM-R scores.

The most intensive duration and data collection schedule studied by Christ et al. (2013c) was collecting one CBM-R score five times per week. If good-quality data is assumed (residual error = 15), monitoring progress daily will provide reliable and valid growth estimates in 12 weeks (low-stakes decisions) and 20 weeks (high-stakes decisions). If poor-quality data are assumed, observed CBM-R scores provide reliable and valid growth estimates in 16 weeks (low-stakes decisions) and did not meet criteria for high-stakes decisions after 20 weeks of progress monitoring (i.e., 100 CBM-R data points). The error in slope estimates was approximately 0.35 WRC/M when data were reliable for low-stakes decisions regardless of data quality. Data did not provide reliable or valid estimates of student growth prior to 6 weeks, even if collected daily. This important finding was corroborated by a recent field-based study (Thornblad and Christ 2014).

Estimates were also collected assuming a moderate data-collection schedule of three observations per day, on 2 days per week. The collection of three data points and using the median score to estimate performance is similar to procedures recommended for universal screening. Results suggested a similar duration of progress monitoring was needed to make low-stakes decisions as when one CBM-R assessment was conducted daily (Christ et al. 2013c). Assuming good-quality data, high-stakes decisions could be made in 18 weeks. The error in slope estimates was approximately 0.29 WRC when using the median of the three scores (Christ et al. 2013c).

When the data collection schedule of three observations on one day was thinned to once per week, the required duration of progress monitoring was extended further. If good-quality data are assumed, CBM-R data provide reliable and valid growth estimates in 14 weeks (low stakes decisions). Assuming poor-quality data, observed CBM-R scores provide reliable and valid growth estimates in 20 weeks (low stakes decisions). High-stakes decisions were not possible using good or poor-quality data sets. High-stakes decisions were possible after 20 weeks assuming very good data, but field-based collection of very good-quality data may not be plausible at this time (Christ et al. 2012). Regression analyses suggested that duration accounted for the most variation when estimating student growth, but the importance of the schedule of data collection and passage set quality was not trivial.

An alternative to collecting weekly progress-monitoring data that has been explored involves estimating growth using a pre/post design. Under this format, student growth is estimated by comparing students' baseline performance to performance after a period of intervention. This would lessen the burden of assessment on teachers (Jenkins and Terjeson 2011) and still capture the effects of the duration of intervention, which made up a large amount of variance in student growth estimates (Christ et al. 2012, 2013c). Indeed, two studies found that slope estimates collected at baseline and after 8 or 10 weeks of instruction were similar to slopes estimated from more frequent data collection (Jenkins et al. 2009; Jenkins and Terjeson 2011). The slope estimates from a pre/post collection schedule had higher sensitivity than most slopes estimated from more frequent collection (Jenkins et al. 2009). These findings are important, but limited by the relatively short duration and the use of weekly CBM-R data as a proxy of true slopes (Christ et al. 2013c). More recent simulation research tested the use of a pre/post method for estimating growth across different durations and data set qualities. Assuming good- or poor-quality data sets, growth estimates did not reach acceptable validity or reliability for low-stakes decisions (Christ et al. 2013c). Growth estimates were reliable and valid for low-stakes decisions after 14 weeks, when data quality was deemed high quality. Again, progress-monitoring data collected in practice seem unlikely to meet this standard.

Implications for Practice Previous recommendations of collecting 10 data points underestimated the amount of data required to estimate growth, at least for CBM-R. More recent research indicated that educators would need to collect quality data on a daily basis for 12 weeks before making a low-stakes decision and 20 weeks for making a high-stakes decision. Thinning the schedule to using median scores twice per week provided similar results. Collecting 1 median score or 1 data point per week would not provide reliable enough data to make high-stakes decisions after 20 weeks of data collection (Christ et al. 2013c). Using pre/post methods did not improve the reliability or validity of growth estimates assuming good- or poor-quality data (Christ et al. 2013c), thus not providing interventionists with any formative data to inform instruction. Based on these findings, interventionists should assess the student's progress daily (using one probe) or twice weekly (using mean scores from three probes). The duration of progress monitoring depends on the types of decisions made (see below).

Making Decisions Based on Progress-Monitoring Data

The purpose of monitoring student progress is to improve the accuracy of decisions and inform more effective instruction. Student progress is measured over time to facilitate data-based decisions. Goals are typically ambitious for students receiving intensive interventions, and decision rules are used to facilitate educators' interpretations of progress-monitoring data in relation to student goals (Shapiro 2011). These rules are an important aspect of monitoring student progress (see Stecker et al. 2005) but recent research calls their accuracy into question.

Recall that most recommendations suggested that 10 data points are needed before making a decision (Ardoin et al. 2013). After these data were collected, educators could make reliable decisions based on the pattern of student performance. Decision rules focus on the use of data points or trend lines (i.e., calculating a slope to predict future performance). Data-based decision rules are typically based on the final three to six scores (e.g., Deno 1986; Hintze and Marcotte 2010; Stecker et al. 2008). If all five points are above the goal line, the intervention is continued and the goal can be increased or intervention intensity can be reduced (i.e., student returned to a lower tier of supplemental support). If the five points are distributed above and below the goal line, the intervention is continued and if all five points are below the line then the intervention is modified or potentially discontinued in favor of more intensive support.

Results from several recent studies suggest that data-point decision rules may lead to inaccurate decisions. The error evident in individual point estimates from well-researched CBM-R measures (e.g., 11–15 WRC/M; Ardoin and Christ 2009) makes rules based on data points particularly worrisome. This point is illustrated in Fig. 1. The graph shows hypothetical data for a third-grade student performing near the tenth percentile using national norms (Hasbrouck and Tindal 2006). An ambitious goal was set (2.0 WRC/M) and progress was monitored one time per week for 10 weeks. The error bars surrounding the observed performance indicate the 68 % confidence interval for the prediction ($SEE = 12$, Ardoin and Christ 2009). Although five consecutive data points fell below the line, considering the error, it is difficult to determine if a change is warranted. Indeed, decisions based on data-point decision rules had very low reliability for students receiving tier 2 interventions (Burns et al. 2010). This error, and the lack of empirical

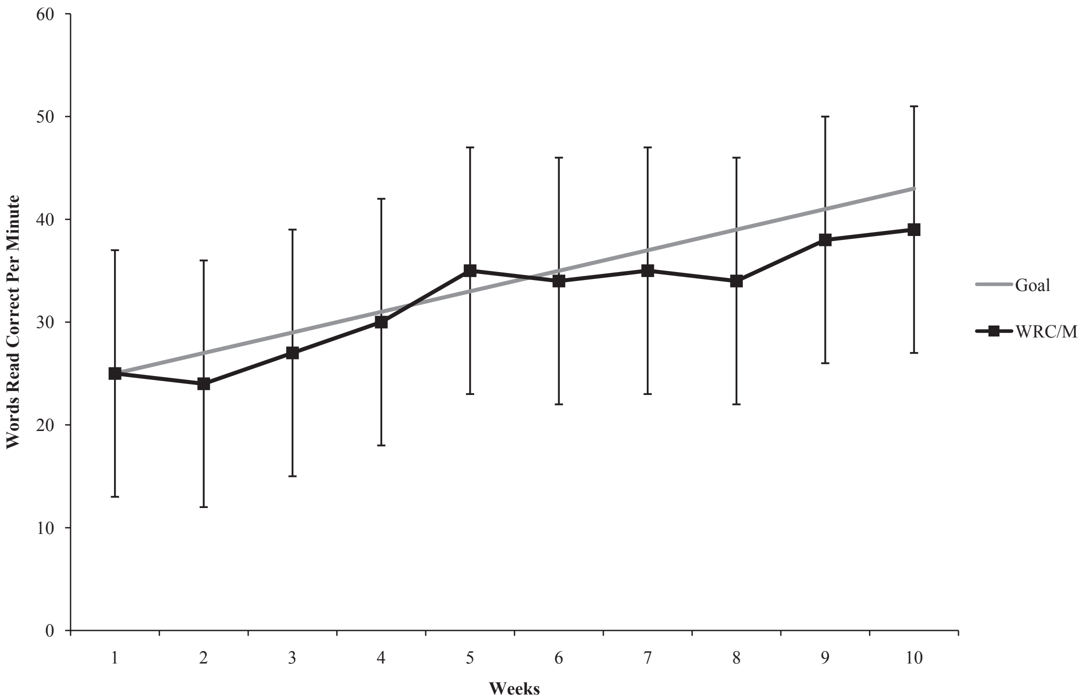


Fig. 1 Hypothetical CBM-R Data for a third-grade student. Data represent 1 probe collected weekly. Error bars represent standard error of the estimate, estimated at 12 WRCM. Goal line represents weekly improvement of 2 WRCM per week. WRCM = Words Read Correct Per Minute

support for data-point rules, indicates educators should discontinue the use of data-point rules to inform instructional decisions until more research establishes their adequacy (cf. Ardoin et al. 2013).

Trend line rules require the use of ordinary least squares regression to estimate the projected rate of growth based on observed data. Calculating slopes can be done with widely available software such as Microsoft Excel. The estimated slope is then compared to the student's goal line. When the observed slope is greater than the goal line the intervention can be continued and the goal is increased or intervention intensity could be reduced (i.e., student moved to a lower tier of supplemental support). When the observed growth is similar to the goal line the intervention is continued. Finally, if the observed slope is below the goal line the intervention is modified or discontinued in favor of providing more intensive support.

Consider the same hypothetical data shown again in Fig. 2. This student's current rate of growth is 1.65, which is below the established goal. The interventionist could decide to modify the intervention or discontinue the intervention based on traditional decision rules. The student's true growth is expected to range from 0.94 to 2.4 WRC/M approximately 68 % of the time, after factoring in the standard error of the slope (0.71 WRC/M; Ardoin and Christ 2009). The lack of precision in the slope estimate makes the decision less clear.

Results from the simulation studies conducted by Christ et al. (2013c) suggest that collecting more data of slightly higher quality (residual error = 10) over the same 10 weeks would make a marginal difference. Note that this is an imperfect comparison due to differences between the error estimates (standard error of slopes, $SEb = 0.71$) observed by Ardoin and Christ (2009) and the error estimates (root-mean-square error $RMSE = 0.49$) simulated in Christ et al. (2013c).

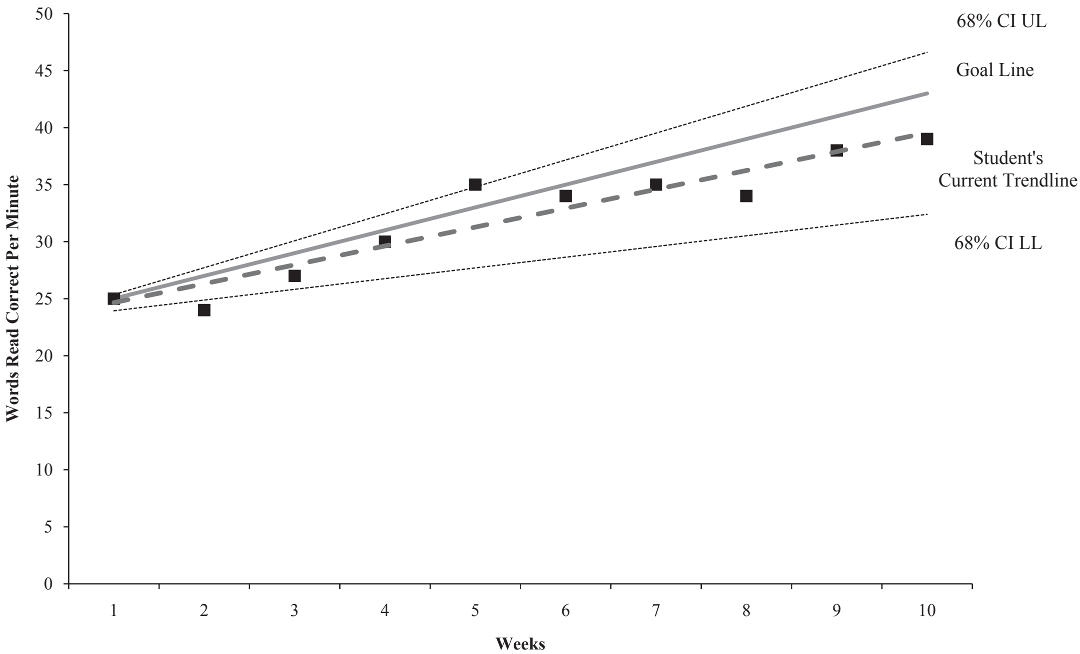


Fig. 2 Hypothetical CBM-R data for a third grade student. Data represent 1 probe collected weekly. Residual error = 12 WRCM. Goal line represents improvement of 2 WRCM per week. Dashed line represents student's current growth (1.65 WRCM per week). Dotted lines represent 68 % CI (*SEb* = 0.71; Ardoin and Christ, 2009). WRCM = Words Read Correct per Minute, CI confidence interval, *SEb* standard error of slopes

However, for a student with an observed rate of growth of 1.65 WRC/M, the true rate of growth is expected to range from 1.16 to 2.14. More importantly, these growth estimates would not meet psychometric criteria for making low-stakes decisions (Salvia et al. 2013).

Findings from Christ et al. (2012, 2013c) suggest that enough data can be collected to provide reliable growth estimates but only under very specific conditions. Our estimation of the student's true performance would likely improve if more data are collected over a longer duration. Assessing the student's progress daily or using the median score of three probes collected twice per week, would provide data adequate for making low-stakes decisions after 12 weeks ($b = 1.65$ WRC/M; 68 % CI = 1.28 to 2.02). Either collection schedule would provide estimates adequate enough for making high-stakes decisions after 20 weeks ($b = 1.7$ WRC/M; 68 % CI = 1.48 to 1.82), at least based on the simulated parameters in Christ et al. (2013c).

Educators are still required to make data-based decisions within RTI systems, despite the problems inherent in decision rules. Validation of decision rules requires examining the expected consequences from an outcome (Kane 2013b). An important decision for students receiving intensive support is whether to continue, modify, or discontinue an intervention. Decisions to modify interventions may include changes in skill targets, materials, or the selection of a different intervention of equal intensity. Incorrect decisions regarding intervention modification may lead to a student receiving an ineffective intervention for a longer period of time, or an effective intervention being replaced with an ineffective one. This is a low-stakes decision because the decision is tested through continued progress monitoring (Riley-Tillman et al. 2013).

Decisions to modify an intervention based on progress-monitoring data are subject to errors inherent in the measures (Ardoin and Christ 2009). Waiting to modify a potentially ineffective intervention for 12–20 weeks is a distressing option. More data are needed to better inform instructional modification (Shapiro 2013).

Mastery measures are useful for improving intervention modification decisions. Mastery data are designed to be more sensitive to skill acquisition than general outcome measures such as CBM-R. Although the additional value of collecting mastery data for informing intervention modification is an empirical question that needs be tested, there is theoretical support for the practice. Student performance on these measures could be analyzed together before determining if instructional modifications are needed (e.g., VanDerHeyden and Burns 2005). These data may provide direction for future modifications as well.

General Implications and Future Directions

Progress monitoring is a core component of RTI systems. At the most intensive level, these data inform decisions regarding short-term instructional modifications and growth over time toward grade-level objectives. Student performance at this level may also be used to inform decisions about special education eligibility. Educational decisions must be made with data that are technically adequate. Technical adequacy differs depending on the proposed uses and interpretations of the data (Kane 2013b).

Recent research has several implications for applied practice (see Table 1). Traditional recommendations underestimated the number of weeks, and number of assessments per week, required to effectively monitor progress. More research is needed to replicate the findings regarding CBM-R precision and also to extend these lines of research into other skill areas. Yet, the plausibility of making correct decisions on 6–10 weeks of data collection appears low once considering the error found in field-based and simulation studies. Educators and researchers alike must take steps to minimize the error in the decisions made about students.

For example, evidence suggests educators should not make decisions based on data-point decision rules. There seems to be too much error in individual point estimates for these to be accurate. It is possible that future research will indicate these rules are suitable for math or other measures but until that exists, decisions should be made differently. Collecting more data over a longer duration of time should also improve the reliability of decisions made based on trend lines. The duration and schedule of data collection necessary are based on a number of factors including error inherent in the measures and data collection procedures. Other sources of error cannot be totally minimized. An exact number of weeks or schedule of data collection is unknown but preliminary estimates could be used as guidance (e.g., Christ et al. 2012, 2013c). The cost–benefit ratio of extending data collection and intensity is unknown but the negative consequences for misclassification seem greater than continuing to provide intensive academic intervention.

Educators can immediately begin recognizing the potential of error when monitoring the progress of students. This can be done by selecting measures with the lowest estimates of standard error, adhering to standardized data collection procedures, and interpreting the results in context of their limitations. For example, AIMsweb now incorporates the use of confidence intervals when graphing student scores (Pearson Inc. 2012). Educators must consider the potential impact of error when visually assessing progress-monitoring graphs.

Future Directions for Research

The directions for future research on individual progress monitoring abound. The use of GOMs was related to dissatisfaction with norm-referenced assessments with limited instructional utility. More evidence is needed to determine the extent to which general outcome measures can be used effectively to monitor individual student progress (Shapiro 2013).

As mentioned, the recent findings regarding CBM-R can only be generalized to that measure. Research is needed to identify the conditions that progress-monitoring data can be used to reliably and validly estimate student growth in reading comprehension, math, spelling, early writing,

Table 1 Implications for Practice

Implications for progress monitoring students receiving intensive support
1. Decisions based on progress-monitoring data range are the often high stakes the most intensive at tier 3. Measures for progress monitoring should have reliability and validity estimates of > .70 for low-stakes decisions and > .90 for high-stakes decisions (see Salvia et al. 2013)
2. Strong evidence supports the use of curriculum-based measures (CBM) for screening and benchmarking purposes. This evidence should not be generalized to the use of general outcome measures such as CBM for making individual decisions
3. Mastery measures and general outcome measures are both needed to effectively monitor progress for students receiving intensive support
4. Mastery measures are suitable for making short-term instructional changes. Mastery measures provide information about the student's acquisition of a narrowly defined skill that is the target of ongoing instruction
5. Mastery measures can be teacher created or commercially developed. Items should be of equal difficulty and provide enough opportunity for the student to demonstrate progress. These measures should provide data regarding accuracy and fluency with a skill (both are required for proficiency). It is also possible to use assessments such as single-skill CBM probes for monitoring subskill mastery
6. General outcome measures are correlated with standardized tests of achievement and state tests. Most measures used today are independent of the curriculum but assess robust skill indicators required for progress in broad skill areas
7. General outcome measures that are commercially developed are typically well constructed and equated for difficulty. However, even the best measures are subject to measurement error. This error varies across measures and the amount of error should be an important consideration for educators when selecting these measures
8. Traditional recommendations for the duration, frequency, and interpretation have been challenged by recent research on CBM for reading (CBM-R) specifically. Results do not generalize to other measures (early literacy, writing, math), but, they do provide reason to reevaluate progress-monitoring practices in schools
9. Duration appears to be a major factor when estimating student progress over time. Previous recommendations suggested 10 weeks of data collection but data at this time are likely unreliable. Estimates based on simulation of CBM-R data suggest progress should be monitored for 14–20 weeks depending on decisions to be made
10. Baseline progress should be estimated using the median score of 3 assessments delivered at 1 time
11. Student progress should be measured more than once per week. Administering three probes and using the median score twice per week or administering one probe daily will likely provide a more reliable estimate of actual progress. It seems reasonable to administer 1 assessment per day of intervention prior to beginning. The influence of more intensive data collection in other skill areas is unknown, but collecting more data will provide more accurate estimates of true growth
12. Data collection procedures must be standardized and assessment data collected with integrity. Research about the impact of departures from standardization is equivocal. However, the high-stakes nature of the decisions made based on progress-monitoring data within tier 3 requires the most data are being 13. Educators should not use data-point rules to make educational decisions for students receiving tier 3 support. These rules are influenced by the error in data-point estimates and also in establishing goal lines. Moreover, these rules were recommended and not based on empirical data (Ardoin et al. 2013)
14. Trend line rules should only be used when enough data can be collected to accurately estimate student growth. More research is needed to establish clear guidelines but the guidelines provided by Christ et al. (2013c) seem to be a minimum starting point
15. When decisions must be made before 14–18 weeks of progress monitoring, they should be made based on multiple data sources and in light of the potential error in the data. The combination of mastery data and GOM data is one option for informing instructional modifications, although it has not yet been empirically tested
16. Educators can choose to monitor progress at grade level or at the student's instructional level. If grade-level monitoring is chosen, it is important to recognize the student will likely show little growth until this becomes the instructional level (Shapiro, 2011). Another option is to monitor progress at the instructional level and collect triennial benchmarking data at grade level to monitor growth toward year-end learning targets
17. Educators must stay up to date with research in this area. As more research on different skill areas continues, there may be important implications for effective progress monitoring

and early literacy skills. This research is not limited to measures similar to CBM. For example, computer adaptive tests seem promising due to the individualization of item administration and the potential for providing both mastery and GOM data (Shapiro 2013).

Research is also needed to determine if progress-monitoring data can be used in applied settings to improve educational outcomes. Formatively evaluating instruction has been associated with improved student achievement (Fuchs et al. 1989; Fuchs et al. 1991) but the associated improvement could also be due to attention to instruction or other factors (Ardoin et al. 2013). More work to provide educators with empirical guidelines for making valid decisions, technology to interpret the impact of measurement error (e.g., calculation of confidence intervals surrounding growth estimates based on national norms), and higher-quality measures or procedures to reduce measurement error to acceptable values will hopefully lead to improved assessment practices in consideration of the increasing demands in education.

Conclusions

Effectively monitoring the progress of students receiving intensive interventions may never have appeared more daunting. General outcome measures designed for instructional modification have been adopted for uses that require strict scrutiny about the reliability and validity of these measures. The proliferation of assessing a student's response-to-intervention was influenced by to dissatisfaction with achievement–ability discrepancy models (Vaughn and Fuchs 2003). Until recently, however, these measures had not been evaluated enough for such purposes. The results discussed throughout this chapter do not suggest that a return to a discrepancy model is needed. Instead, more research is needed to determine how to properly assess a student's response-to-intervention effectively. Educators should continue to make decisions based on multiple sources of data while acknowledging the limitations inherent within the measures used. Although some may question the use of CBA data within RTI systems, it seems entirely too early to make any definitive conclusions. More data are needed before a reliable decision can be made.

References

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from Dibels, Aimsweb, and an Experimental passage set. *School Psychology Review*, *38*, 266–283.

Ardoin, S. P., Christ, T. J., Moreno, L., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of recommendations and research surrounding curriculum based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, *51*, 1–18. doi:10.1016/j.jsp.2012.9.004.

Burns, M. K. (2001). Measuring sight-word acquisition and retention rates with curriculum-based assessment. *Journal of Psychoeducational Assessment*, *19*, 148–157. doi:10.1177/073428290101900204.

Burns, M. K. (2004). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education, 25,* 167–175. doi:10.1177/07419325040250030401.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School Psychology Quarterly, 22,* 297–313. doi: 10.1037/1045-3830.22.3.297.

Burns, M. K., & Gibbons, K. (2012). *Response to intervention implementation in elementary and secondary schools: Procedures to assure scientific-based practices* (2nd ed.). New York: Routledge.

Burns, M. K., & Klingbeil, D. A. (2010). Assessment of academic skills in math within a problem-solving model. In G. G. Peacock, R. A. Ervin, K. W. Merrell, & E. J. Daly (Eds.), *Practical handbook of school psychology: Effective practices for the 21st century* (pp. 86–98). New York: Guildford.

Burns, M. K., & Mosack, J. L. (2005). Criterion-related validity of measuring sight-word acquisition with curriculum-based assessment. *Journal of Psychoeducational Assessment*, *23*, 216–224. doi:10.1177/073428290502300302.

Burns, M. K., & Parker, D. C. (2014). *Curriculum-based assessment for instructional design: Using data to individual instruction*. New York: Guilford.

Burns, M. K., Tucker, J. A., Frame, J., Foley, S., & Hauser, A. (2000). Interscorer, alternate-form, internal consistency, and test-retest reliability of Gickling's model of curriculum-based assessment for reading. *Journal of Psychoeducational Assessment*, *18*, 353–360. doi:10.1177/073428290001800405.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. L. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review*, *35*, 401–418.

Burns, M. K., Scholin, S. E., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment*, *28*, 102–114. doi:10.1177/0734282909342374.

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimates of standard error of the slope to construct confidence intervals. *School Psychology Review*, *35*, 128–133.

Christ, T. J., & Vining, O. (2006). Curriculum-based measurement procedures to develop multiple-skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements. *School Psychology Review*, *35*, 387–400.

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children*, *78*, 356–373.

Christ, T. J., Monaghen, B. D., Zopluoglu, C., & Van Norman, E. R. (2013a). Curriculum-based measurement of oral reading: Evaluation of growth estimates derived with pre-post assessment methods. *Assessment for Effective Intervention*, *38*, 139–153. doi:10.1177/1534508412456417.

Christ, T. J., White, M. J., Ardoin, S., & Eckert, T. L. (2013b). Curriculum-based measurement of reading: Consistency and validity across best, fastest, and question reading conditions. *School Psychology Review*, *42*, 415–436.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. R. (2013c). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology*, *51*, 19–57. doi:10.1016/j.jsp.2012.11.001.

Colón, E. P., & Kranzler, J. H. (2006). Effect of instructions on curriculum-based measurement of reading. *Journal of Psychoeducational Assessment*, *24*, 318–328. doi:10.1177/0734282906287830.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*, 394–409. doi:10.1037/0022-0663.98.2.394.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, *102*, 327–340. doi:10.1037/a0018448.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219–232.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review*, *15*, 358–374.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*, 184–192. doi:10.1177/00224669030370030801.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, *30*, 507–524.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics. A review of the literature. *The Journal of Special Education*, *41*, 121–139. doi:10.1177/00224669070410020101.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, *46*, 315–342.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–192.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, *57*, 488–499.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children*, *61*, 15–24.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*, 199–208.

Fuchs, D., & Fuchs, L. S. (2011). *Using CBM for progress monitoring in reading*. Washington, DC: National Center on Student Progress Monitoring.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, *21*, 449–460. doi: 0.3102/00028312021002449.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *The Journal of Educational Research*, 103–110. doi:10.2307/1163151.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Education Research Journal*, *28*, 617–641. doi:10.3102/00028312028003617.

Gickling, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprehension. *Journal of Learning Disabilities*, *11*, 559–566. doi:10.1177/002221947801100905.

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children*, *52*, 205–218.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*, 636–644.

Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review*, *33*, 204–217.

Hintze, J. M., & Marcotte, A. M. (2010). Student assessment and data-based decision making. In T. A. Glover & S. Vaughn (Eds.)*, The promise of response to intervention* (pp. 57–77). New York: Guilford.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*, 372–386.

Hintze, J. M., Shapiro, E. S., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education*, *28*, 188–202.

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review*, *31*, 514–528.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools*, *43*, 45–56. doi:10.1002/pits.20128.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade. *School Psychology Review*, *34*, 9–26.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2008). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.

Hosp, J. L., Hosp, M. K., Howell, K. W., & Allison, R. (2014). *ABCs of curriculum-based evaluation: A practical guide to effective decision making*. New York: Guilford Press.

Jenkins, J., & Terjeson, K. J. (2011). Monitoring reading growth: Goal setting, measurement frequency, and methods of evaluation. *Learning Disabilities Research & Practice*, *26*, 28–35. doi:10.1111/j.1540-5826.2010.00322.x.

Jenkins, J. R., Graff, J. J., & Miglioretti, D. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children*, *75*, 151–163.

Kane, M. (2013a). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73. doi:10.1111/jedm.12007.

Kane, M. (2013b). The argument-based approach to validation. *School Psychology Review*, *42*, 448–457.

Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review*, *37*, 374–390.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why to do it. In M. R. Shinn (Ed.)*, Curriculum-based measurement: Assessing special children* (pp. 18–44). New York: Guilford.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education*, *41*, 68–84. doi:10.1177/00224669070410020301.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice*, *24*, 186–195. doi:10.1111/j.1540-5826.2009.00292.x.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*, 5–11. doi:10.3102/0013189X018002005.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749. doi:10.1037/0003-066X.50.9.741.

National Center on Intensive Intervention. (2013). *Data-based individualization: A framework for intensive intervention*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, Author.

National Center on Response to Intervention. (2010). *Essential components of RTI—A closer look at response to intervention*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, Author.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. http://www.nichd.nih.gov/publications/nrp/report.cfm. Accessed 1 July 2014.

Parker, D. C., McMaster, K. L., & Burns, M. K. (2011). Determining an instructional level for early writing skills. *School Psychology Review*, *40*, 158–167.

Pearson Inc. (2012). *Aimsweb: Progress monitoring guide*. Bloomington: Author.

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum- based measurement. *Journal of Psychoeducational Assessment*, *23*, 326–338. doi:10.1177/073428290502300403.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, *47*, 427–469. doi:10.1016/j.jsp.2009.07.001.

Riley-Tillman, T. C., Burns, M. K., & Gibbons, K. (2013). *RTI applications, Volume 2: Assessment, analysis, and decision making*. New York: Guilford.

Salvia, J., Ysseldyke, J., & Bolt, S. (2013). *Assessment in special and inclusive education* (12th ed.). Belmont: Wadsworth Cengage.

Shapiro, E. S. (2008). Best practices in setting progress monitoring goals for academic skills improvement. In A. Thomas & J. Grimes (Eds.)*, Best Practices in School Psychology* (pp. 141–157). Bethesda: National Association of School Psychologists.

Shapiro, E. S. (2011). *Academic skills problems* (4th ed.). New York: Guilford.

Shapiro, E. S. (2013). Commentary on progress monitoring with CBM-R and decision making: Problems found and looking for solutions. *Journal of School Psychology*, *51*, 59–66. doi:10.1016/j.jsp.2012.11.003.

Shapiro, E. S., Keller, M. A., Lutz, J. L., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests. *Journal of Psychoeducational Assessment*, *24*, 19–35. doi: 10.1177/0734282905285237.

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, *30*, 407–419.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice*, *15*, 128–134. doi:10.1207/SLDRP1503_2.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*, 795–819. doi:10.1002/pits.20113.

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision-making. *Preventing School Failure: Alternative Education for Children and Youth*, *52*, 48–58. doi:10.3200/PSFL.52.2.48-58.

Taylor, C. D., Meisinger, E. B., & Floyd, R. G. (2013). Variations in directions and overt timing on oral reading accuracy, fluency, and prosody. *School Psychology Review*, *42*, 437–447.

Thornblad, S. C., & Christ, T. J. (2014). Curriculum-based measurement of reading: Is 6 weeks of daily progress monitoring enough? *School Psychology Review*, *43*, 19–29.

Treptow, M. A., Burns, M. K., & McComas, J. J. (2007). Reading at the frustration, instructional, and independent levels: The effects on students' reading comprehension and time on task. *School Psychology Review*, *36*, 159–166.

Tucker, J. A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children*, *52*, 199–204.

University of Oregon Center on Teaching and Learning. (2012). *DIBELS Next recommended benchmark goals*. Eugene: University of Oregon.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention*, *30*, 15–31. doi:10.1177/073724770503000302.

Van Norman, E. R., Christ, T. J., & Zopluoglu, C. (2013). The effects of baseline estimation on the precision of CBM-R growth estimates. *School Psychology Quarterly*, *28*, 239–255. doi:10.1037/spq0000023.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*, *18*, 137–146. doi:10.1111/1540-5826.00070.

Vaughn, S., & Fuchs, L. S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*, 195–203. doi:10.1177/0022219412442150.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measures in reading. *The Journal of Special Education*, *41*, 85–120. doi:10.1177/0022466907 0410020401.

Yeo, S., Kim, D., Branum-Martin, L., Wayman, M. M., & Espin, C. A. (2012). Assessing the reliability of curriculum-based measurement: An application of latent growth modeling. *Journal of School Psychology*, *50*, 275–292. doi:10.1016/j.jsp.2011.09.002.

Ysseldyke, J., Burns, M. K., Scholin, S. E., & Parker, D. C. (2010). Instructionally valid assessment within response to intervention. *Teaching Exceptional Children*, *42*, 54–61.

Introduction to Problem Analysis to Identify Tier 3 Interventions: Brief Experimental Analysis of Academic Problems

Melissa Coolong-Chaffin and Jennifer J. McComas

In light of recent policy shifts, many school systems across the nation are adopting multi-tiered systems of support (also known as response to intervention or RTI) in order to meet the academic needs of all students (National Center on Response to Intervention (NCRTI) 2013). An RTI framework seeks to meet the needs of all students by allocating resources according to the intensity of the problem. In general, quality core instruction is provided to all students and students' skills are screened at three benchmark points across the year in tier 1. For students not meeting benchmark standards, "targeted" supplemental interventions or supports are added to their programming within tier 2. Students' progress in tier 2 is monitored more frequently and data are used to determine whether level of performance and/or growth is meeting benchmark standards. For students whose performance is still not adequate after a period of time, more intensive, individualized interventions may be added within tier 3. Across all tiers, care should be taken to ensure that instructional programs, interventions, and assessment tools selected meet evidence standards to the extent possible and that all practices are conducted with high levels of fidelity (NCRTI 2013).

Researchers and practitioners have used different criteria to identify which students are in need of targeted and intensive interventions as well as what those targeted and intensive services should look like. However, decisions about who needs additional resources, what resources are needed, and when changes should be made are all driven by data (NCRTI 2013). This is generally consistent with the problem-solving approach described by Kratochwill and Bergan (1990), whereby practitioners apply a standard set of stages and questions to address student issues. The stages include problem identification (in which data are collected to define the problem), problem analysis (in which data are analyzed to understand why it is occurring and a plan is developed), plan implementation (in which the intervention is delivered, and implementation integrity and progress monitoring data are collected), and plan evaluation (in which data are reviewed to determine whether the problem is still occurring, and whether the plan should be changed or discontinued).

Within a general problem-solving framework, two different approaches have been used to select targeted and intensive interventions. These approaches have been described as RTI-standard protocol and RTI-problem-solving (Fuchs et al. 2003) also known as RTI-problem analysis (Christ et al. 2005), the term which the authors use here. The RTI-standard protocol approach typically involves applying the same empirically validated protocol to all students who struggle with a given academic skill. In contrast,

M. Coolong-Chaffin (✉)
University of Wisconsin, Eau Claire, USA
e-mail: chaffimc@uwec.edu

J. J. McComas
University of Minnesota, Minneapolis, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_21

the RTI-problem-solving or analysis approach analyzes contextual variables (e.g., instruction, environment, curriculum) as well as learner variables in order to identify interventions that "fit" the unique needs of the student (see Tilly 2008).

Some have argued that in order for interventions to be truly "individualized" it is not enough to simply provide interventions in a small or 1:1 setting. Rather, tier 3 interventions must be matched to the unique characteristics of learners with the most intense needs to be effective (Burns and Coolong-Chaffin 2006). The purpose of this chapter is to describe an experimental approach that can be used to select individualized interventions for students who have not made adequate progress given a more general approach to supplemental intervention in tier 2.

An Experimental Approach to Selecting Academic Interventions

Attempts to systematically identify interventions that are effective for improving academic-related behavior are not new. As early as the very first issue of the *Journal of Applied Behavior Analysis* in 1968, the effects of instructional variables on academic behavior have been examined via systematic manipulation. Hart and Risley (1968) sought to increase the use of adjectives in everyday language of preschool children and found a functional relationship between contingent access to materials and adjective use in spontaneous language. Contingent reinforcement has also been used for several decades to improve reading performance (Gray et al. 1969; Whitlock 1966). However, the early studies of the effect of reinforcement on reading performance were arguably designed to demonstrate the viability of reinforcement as an intervention for particular populations of students rather than identifying the specific learning situations and procedures that produced desired effects for a given student (Bond and Dykstra 1967). Nonetheless, these early studies of intervention effects on reading performance were based on the following assumptions (Gray et al. 1969) which still hold true today:

1. Reading is a learning task; therefore, it is responsive to the laws of learning
2. Decoding is antecedent to comprehension and therefore, should be attended to first
3. The system should provide success in abundant quantities and should provide frequent and immediate reward for that success
4. Characteristics such as visual perception, intelligence quotient (IQ), visual–motor coordination, visual and auditory sequencing, etc. are considered not to affect differentially prognosis for skill mastery unless the deficiency is of such a marked and gross degree that it results in reduced functional ability in nonreading activities.

The assumption that the laws of learning govern reading performance of individual students appeared to gain traction in the 1990s. Researchers considered principles of learning such as stimulus control and examined the relative effects of multiple instructional strategies for one individual to select instructional procedures or strategies based on the relative responsiveness to them for reading comprehension (see Fig. 1; McComas et al. 1996), spelling (see Fig. 2; McComas et al. 1996), and handwriting (see Fig. 3; McComas 1994).

The comparisons of instructional strategies were referred to as functional analyses or functional analyses of reading (FRA; Wagner et al. 2006) because they were designed to demonstrate a consistent effect on a dependent variable by systematically manipulating an independent variable (Kennedy 2005). These brief comparisons of instructional strategies are based on brief functional analyses that were originally designed to specify the operant function (i.e., reinforcer) for a target challenging behavior (e.g., self-injurious behavior). Brief functional analysis procedures were designed to be contingency driven (i.e., the condition introduced in a session was contingent on the effect of the previous condition on the target behavior), to be conducted over a short period of time (i.e., more than 3 but fewer than 12 sessions), and to preserve the principles of "predict and control," in other words, replication of effects within the experimental design. (e.g., Northup et al. 1991). To highlight the

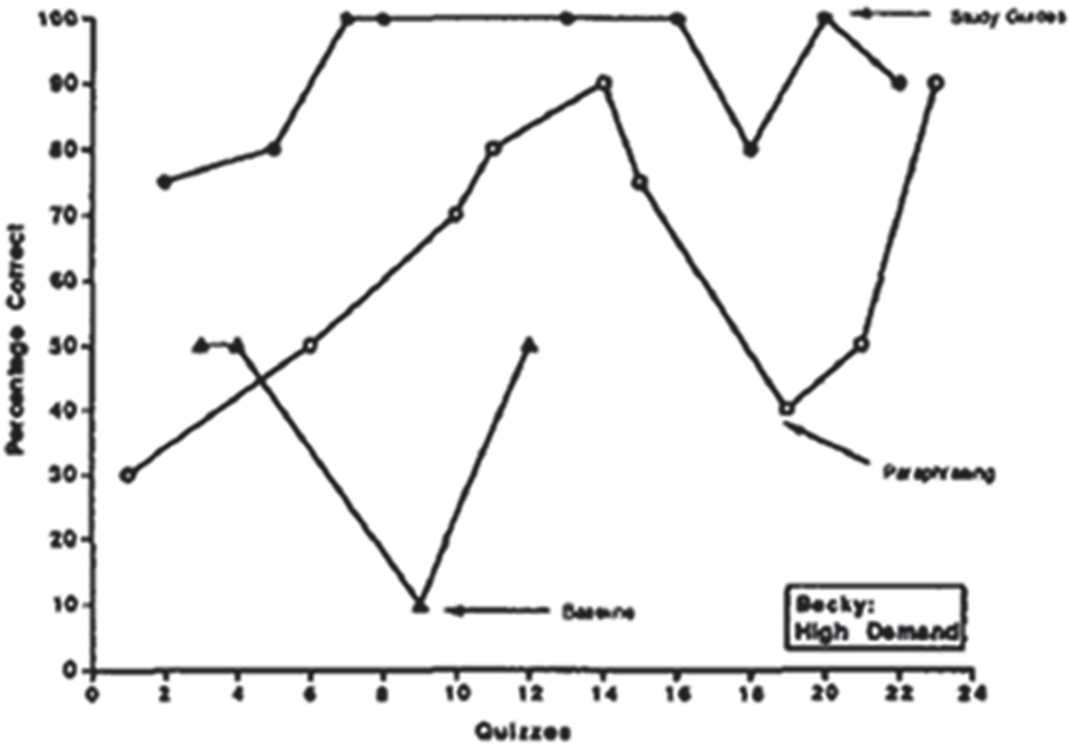


Fig. 1 Effects of paraphrasing and study guides on correct responding on chapter comprehension quizzes of a tenth-grade student diagnosed with learning disabilities. (McComas et al. 1996)

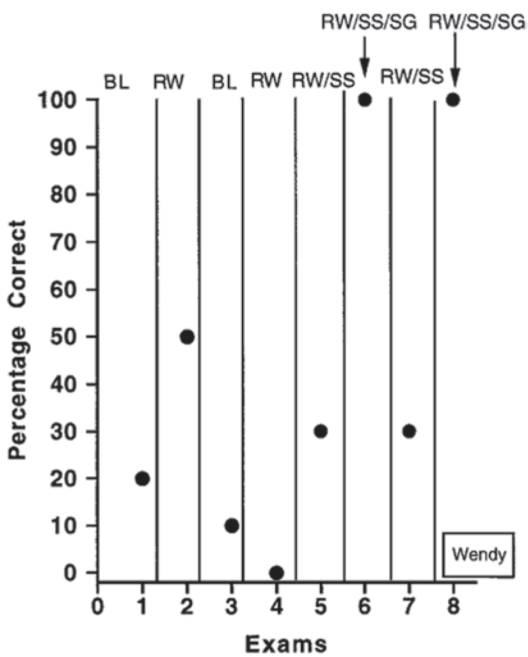


Fig. 2 Effects of rhyming words (RW), rhyming words with sample spelling (RW/SS), and rhyming words with sample spelling of a student-generated word (RW/SS/SG) on correct spelling words. (McComas et al. 1996)

experimental analysis aspect of these procedures and to reduce confusion about the nature of the functional relations under examination (operant function vs. effect of and instructional strategies on academic responding), the term brief experimental analysis (BEA) has been adopted for use to describe systematic analysis of instructional strategies to improve an individual's academic performance.

BEA is an application of single-case experimental design logic. Put simply, "Single-case experimental designs are used to demonstrate experimental control within a single participant" (p. 12, Kennedy 2005). Experimental conditions are tightly controlled with one independent variable being introduced at a time to test its effect on the dependent variable. Experimental control is demonstrated when a change in level or trend is seen in the dependent variable after the introduction of the independent variable via visual analysis of the graphed performance. In BEA, each treatment or intervention is tested rapidly in succession in order to determine its effectiveness relative to

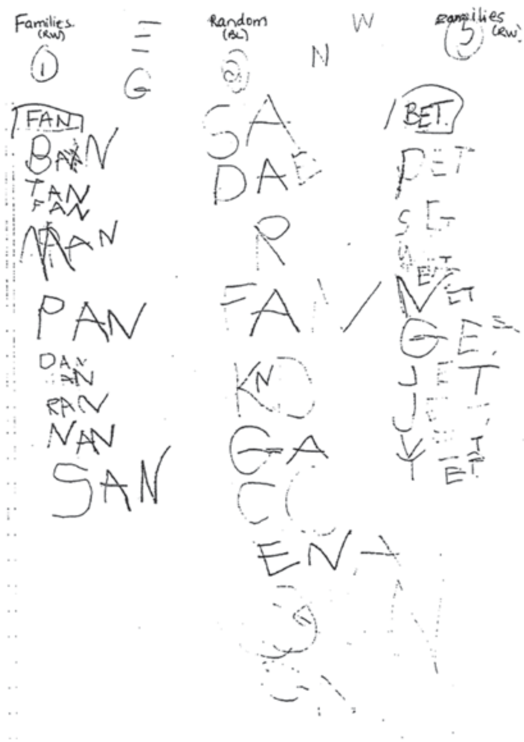


Fig. 3 Effects of rhyming words, rhyming words with sample spelling, and rhyming words with sample spelling of a student-generated word on handwriting. (McComas 1994)

baseline or other interventions (Kennedy 2005). This technique is useful for selecting an effective intervention when time is limited. The selected intervention can then be applied over time and student progress can be measured to determine its effectiveness over a longer time period.

Daly et al. (1997) expanded the use of BEA for choosing and evaluating academic interventions based on a functional explanation of the deficits. Each intervention is designed to test one of the following hypotheses: (a) the child does not want to do the task, (b) the child has not had enough practice to do the task, (c) the child has not had enough help to do the task, (d) the child has not had to do it that way before, or (e) the task is too difficult. By manipulating each independent variable successively through techniques such as incentive, increased practice, modeling, feedback, and task difficulty, while measuring the same dependent variable each time and then replicating the results, the most successful intervention can be selected for each student. The hypotheses are arranged in ascending order from least intrusive to most intrusive, and when tested in that succession, allow the interventionist to determine the most simple, effective intervention for the student.

Basic Procedures

A BEA of academic performance involves these essential components: (a) a measure that is directly related to the skill targeted by the intervention, is independent of other skills, and is sufficiently sensitive to detect immediate effects, (b) at least one hypothesis-based intervention that is circumscribed and can be completed within one brief (e.g., 20 min) instructional period, and (c) a distinct "contrast condition" for use in the within-subject experimental analysis. For example, if the target skill is reading fluency, the measure might be the number of words read correct per min (see Fig. 4). Alternately, if there is a concern about accurate decoding, the measure might be the number of errors made per min or the percentage of words read accurately. If the target skill is letter–sound correspondence, the measure should be specific to the letters targeted during the experimental session, not the full range of 26 letters because a probe that contains all 26 letters will not be very sensitive to the immediate effects of 20 min of intervention focused on one or two letters. Taken together, if the immediate effects of the test condition, or hypothesis-based strategy, are positive, the intervention would be deemed promising for affecting generalized effects over a longer span of implementation.

To illustrate, Petursdottir et al. (2009) measured the effect of two strategies: modeling and goal setting with incentive on young children's performance of letter sounds in a BEA using a letter–sound subskill measure that contained 56 letters with the same proportion of two known and two unknown sounds (top panel, Fig. 5). Their findings indicated, in part, that the subskill measure yielded clear results in the brief analysis and was more sensitive than the more general let-

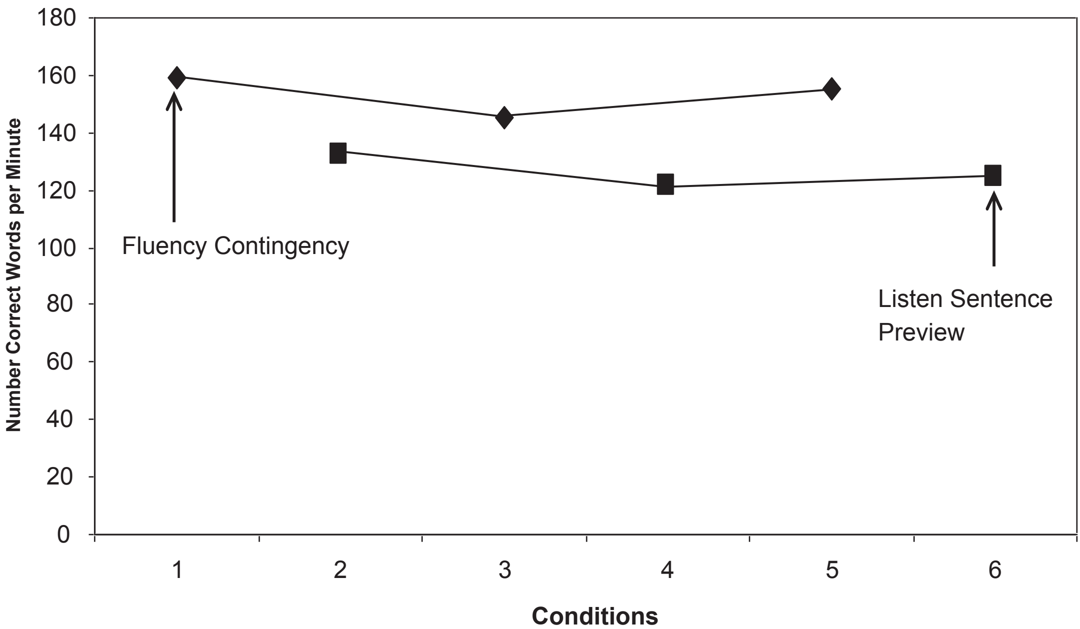


Fig. 4 Number of words read correct during incentive (fluency contingency) and modeling (listen sentence preview) conditions of a brief experimental analysis of reading performance

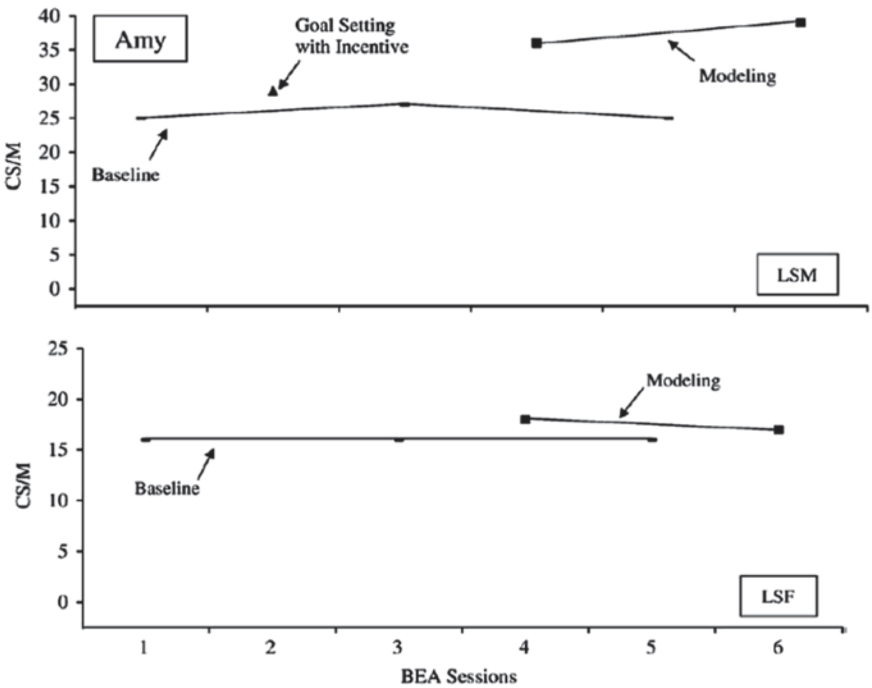


Fig. 5 Illustration of two measures during the same sessions, one more sensitive (letter–sound mastery (*LSM*), *top* panel) than the other (letter–sound fluency (*LSF*), *bottom* panel) during a BEA (Copied with permission from Petursdottir et al. 2009). Both panels depict the number of correct letter sounds per min (CS/M)

ter–sound fluency measure (bottom panel, Fig. 5) that contained all 26 letters of the alphabet.

After the measure is selected, procedures typically involve the following: (a) a pre-intervention measure, (b) intervention, (c) a post-intervention measure, (d) alternation of at least two hypothesis-based test conditions, and (e) replication. Each condition is typically conducted between one and three times and each session lasts less than 30 min, which is why they are considered brief. The effect of the conditions on the academic behavior yields a picture of the functional relationship between the condition (i.e., the instructional strategy) and the academic performance (e.g., number of words read correctly). Manipulating the test conditions via alternating with replication constitutes the analysis.

The results of the process are illustrated in Fig. 4. The instructor was testing the effects of an incentive, which aligns with the hypothesis that the student does not want to do the task, and of modeling, which is the hypothesis that the student has not had enough help to do the task, on the number of words read correctly per minute (WRCM). She began by gathering several different standardized passages so that the passages are of approximately equal difficulty (e.g., at the student's grade level). The instructor selected a passage and conducted a 1-min reading probe with the student on that passage to determine the pre-intervention number of words read correct. Next, the instructor implemented one of the interventions. In this example, the first intervention implemented was a fluency contingency (incentive—student does not want to do it). The instructor told the student that if he read at least 154 words correctly with three or fewer errors, he would be permitted to select a prize from the prize box. Then the instructor conducted another 1-min reading probe on the same passage to determine the effect of the promised prize, or fluency contingency, on the number of words he read correctly.

The instructor then selected a different passage for use in the listen sentence preview (modeling—student needs more help), which was the second intervention. The instructor began the next condition by conducting a 1-min reading probe with the student to determine the pre-intervention number of words read correctly. Next, the instructor implemented the intervention. In this case, the instructor modeled reading at the sentence level. She read a sentence and then instructed the student to read the same sentence back to her. Then the instructor read the next sentence and instructed the student to read back to her the previous two sentences. This requirement to read the previous two sentences continued until they had read approximately 200 words. Then the instructor conducted a post-intervention 1-min reading probe to determine the effect of modeling on the student's number of correct words read per min. The procedures for these two conditions were repeated in alternating fashion until either: (a) the data suggested that at least one intervention demonstrated promising effects or (b) neither intervention had a sufficient effect and therefore an additional intervention was tested. In this case, the fluency contingency was more effective for improving the number of words read correctly, so that intervention was recommended for implementation.

An example of when the first intervention(s) tested are not sufficient appears in Fig. 6. The analysis was conducted with a second-grade girl and began with testing four different conditions using the number of decodable words read correctly in 1 min as the dependent measure. In the first condition, the interventionist modeled how to read each word by saying the sounds in the word then reading the whole word and then the student said the sounds and read each word. The second condition was a repeated reading intervention in which the student read a list of words three times and errors were corrected. The third condition used a phonics song to teach the student how to sound out words. The fourth condition used a "sing it/read it" procedure in which the interventionist modeled reading the sounds in each word, sang the word to model blending the sounds together, then read the word. The student repeated this for each word. None of these conditions resulted in a significant increase in WRCM relative to baseline performance. Thus, an incentive for increased performance was added to the most effective skill-based condition (i.e., sing it/read it), and this condition was compared directly

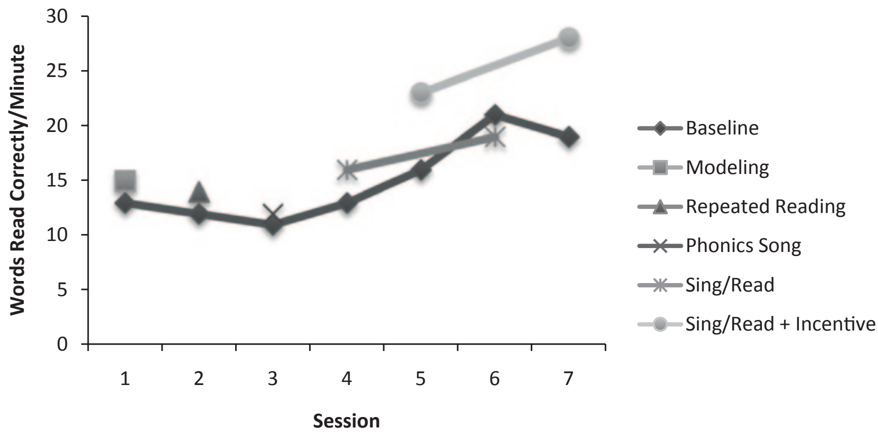


Fig. 6 Four skill-based interventions were initially alternated to examine their effects on words read correctly per minute (WRCM). None of the interventions produced a significant increase in WRCM, so an incentive for increased performance was added to the top skill-based intervention (i.e., sing it/read it). This package was compared to the skill-based intervention alone in sessions 4–7. The combination of sing it/read it plus an incentive for increased performance resulted in the greatest gains and was recommended for use over an extended period of time

to the sing it/read it condition alone across the final four sessions. The student's performance was better in the sing it/read it plus incentive condition, therefore that intervention was recommended for adoption and ongoing progress monitoring.

Evidence for Use

To date, BEA procedures have been used across a variety of academic skills including oral reading fluency, early literacy skills, math computation, and early writing skills. While the vast majority of work has examined oral reading fluency, limited work in the other areas demonstrates the procedures are adaptable and may be useful for other skill areas. Because BEA is an individualized approach based on the logic and principles of single-case design research, it is especially useful in applied settings where practitioners are interested in identifying potentially effective interventions for individual students.

Oral Reading Fluency

Over the past several years, researchers have used BEA techniques to select effective oral reading fluency interventions using connected text for students with and without diagnosed disabilities. Interventions have focused on antecedent variables such as modeling (e.g., Eckert et al. 2003), task difficulty (e.g., Jones and Wickstrom 2002), extra practice (Eckert et al. 2003), and choice of type of instruction (Daly et al. 2006). Other studies have examined the effects of consequence interventions such as offering an incentive for increased performance (e.g., Daly et al. 2005), as well as combinations of antecedent and consequence-based interventions (e.g., Daly et al. 2005; Jones and Wickstrom 2002). Table 1 summarizes interventions that have been described in the literature. In a meta-analytic review of the research on BEA for oral reading fluency, Burns and Wagner (2008) found BEA interventions resulted in an average increase of 30 WRCM over baseline with a mean percent of nonoverlapping data of approximately 82 %, indicating BEA is effective for selecting effective oral reading fluency interventions.

One study examined the use of BEA with English language learners and found that identified interventions improved both oral reading fluency and maze fluency over time for participants (Malloy et al. 2007). In addition, two studies have found parents can implement the interventions effectively as well (Persampieri et al. 2006; Gortmaker et al. 2007).

Table 1 Oral reading fluency interventions used in brief experimental analysis. (Table adapted from Burns and Wagner 2008)

Intervention	Description	Citations
Incentive for increased fluency	The examiner provides an incentive for meeting a predetermined fluency goal. Addresses hypothesis that the student does not want to do the task	Daly et al. 1998, 1999; Eckert et al. 2000, 2003; Jones and Wickstrom 2002; Noell et al. 1998, 2001; Wilbur and Cushman 2006
Incentive for tracking	The examiner provides an incentive for pointing to each word while the student reads aloud. Addresses hypothesis that the student does not want to do the task	McComas et al. 2009
Performance feedback	The examiner tells the student the number of words read correctly and errors. Performance may be displayed on a graph. Addresses hypothesis that the student needs more help with the task	Bonfiglio et al. 2004; Eckert et al. 2000, 2003
Listening passage preview	The examiner reads the passage aloud while the student follows along silently, pointing to each word as it is read. Addresses hypothesis that the student needs more help with the task	Daly and Martens 1994; Daly et al. 1998, 1999, 2002, 2005; Eckert et al. 2000, 2003; Jones and Wickstrom 2002; Noell et al. 1998, 2001; Vanauken et al. 2002
Listening sentence preview	The examiner reads one sentence aloud at a time while the student follows along silently, pointing to each word as it is read. Then the student reads that sentence aloud. The process is repeated until the entire passage is read. Addresses hypothesis that the student needs more help with the task	McComas et al. (2009)
Listening word preview	The examiner reads one word aloud at a time while the student follows along silently, pointing to each word as it is read. Then the student reads that word aloud. The process is repeated until the entire passage is read. Addresses hypothesis that the student needs more help with the task	McComas et al. (2009)
Student passage preview	The student reads the passage silently, and asks the examiner to read unknown words. The examiner reads the unknown words. Addresses hypothesis that the student needs more help with the task	Daly and Martens 1994; Eckert et al. 2003; VanAuken et al. 2002; Wilbur and Cushman 2006
Repeated reading	The student reads the same passage aloud multiple times. Addresses hypothesis that the student needs more practice with the task	Bonfiglio et al. 2004; Daly and Martens 1994; Daly et al. 1998, 1999, 2002, 2005; Jones and Wickstrom 2002; Noell et al. 1998, 2001; Wilbur and Cushman 2006
Phrase drill	After the student reads the passage aloud, the examiner reviews errors. The student reads each word and the sentence or phrase containing the word multiple times. Addresses hypothesis that the student needs more help with the task	Bonfiglio et al. 2004; Daly et al. 1998, 2005; Jones and Wickstrom 2002
Keywords	The student reads a passage aloud and examiner corrects errors. The student identifies up to five words s/he cannot define. The examiner writes each word on a whiteboard, reads it, and asks the student to read the word. Then the examiner defines each word, and uses it in a sentence. Addresses hypothesis that the student needs more help with the task	Malloy et al. 2007
Choice	The student chooses whether s/he would receive instruction prior to reading a passage, how long instruction would last, and what instructional strategy would be used (i.e., modeling, practice, error correction, and performance feedback). Addresses hypothesis that the student does not want to do the task	Daly et al. 2006

Table 1 (continued)

Intervention	Description	Citations
Incremental rehearsal	The same procedure described in keywords is used. This time, the student reads, defines, and uses each word in a sentence. The student practices both known and unknown words in an alternating sequence until all new words are reviewed. Addresses hypothesis that the student needs more help with the task	Malloy et al. 2007
Easier material	Material with readability one grade level lower than actual grade placement is administered. Addresses hypothesis that the task is too hard	Jones et al. 2009

Early Literacy Skills To date, one study has been published regarding the use of BEA procedures with early literacy skills (Pettursdottir et al. 2009). As described previously, these researchers used interventions commonly used for oral reading fluency (i.e., goal setting with incentive, modeling, and a combination of the two) and applied them to teaching letter sounds and high-frequency words for kindergartners with skill deficits. Results indicated the interventions had differentiated results for each participant and that implementation of the BEA-indicated intervention over time yielded increases in skills across participants. These results indicate further research in the area of early literacy skills is warranted.

Math Computation Fewer studies have demonstrated the efficacy of BEA for identifying effective interventions in math. The issue of measures in math was addressed by Vanderheyden and Burns (2009) who demonstrated that skills targeted for intervention could be progress monitored and that performance on progress monitoring probes was positively related to general outcome measures of math computational skills that students are expected to acquire by the end of the school year. Codding et al. (2009) tested two of the Daly et al. (1997) hypotheses: "The child does not want to do the task" and "The child has not had enough practice to do the task" with four participating elementary age students. Experimenters tested the effects of an incentive or goal setting (does not want to do it) for improved digits correct per minute (DCPM) in one condition and a cover, copy, compare (CCC; has not had enough practice to the task) strategy in another condition during 2-min probes using worksheets. Probes contained skills individually identified as necessary targets for each student, ranging from single digit addition for two participants to single digit division for one participant. Incentive was the most effective intervention for one participant, CCC was the most effective intervention for another participant, and for two participants, goal setting was most effective for improving DCPM. An extended analysis demonstrated increases in DCPM with implementation of the BEA-identified individualized intervention with each participant. It should be noted that limited generalized improvement in math skills was observed. The findings were one of the first demonstrations of a BEA with math skills and suggested the need for additional research on the treatment utility of BEA for producing generalized gains in mathematics computation. Replication is needed of BEA with students who are struggling in math to identify not only which intervention is most effective for improving target math skills but also general outcome measures in math.

Writing If few studies have been conducted on the utility of BEA in math, fewer have been conducted in the area of writing. Burns, Ganuza, and London (2009) tested the effects of three interventions on the percentage of target letters formed correctly for a second-grade student who was struggling in the area of written language. The results of the BEA indicated that the modeling strategy was superior to the other instructional strategy and an incentive. An extended analysis of the effects of the modeling strategy demonstrated an increase in accurate formation of all target letters. It should be noted that the observed effects only maintained for one of the three sets of target letters; thus, additional research is needed to examine the treatment utility of BEA-identified interventions for maintenance of skill acquisition in the area of correctly formed letters as well as other aspects of the multidimensional task of writing.

Another aspect of writing is measured in terms of correct letter sequences (CLS). Parker et al. (2012) tested the effects of a performance contingency and three instructional strategies on CLS of three first-grade students who were struggling with writing performance. The instructional strategies included modeling, repeated writing with feedback, and student self-selection of writing stimuli. The BEA identified a different effective intervention for each of the participants and an extended analysis demonstrated continued improvement of CLS with extended implementation of the individualized instructional strategy. Replication of BEA in the area of writing is necessary to more fully evaluate the treatment utility for various aspects of the multidimensional task of writing.

BEA in Practice

Not much is known about the use of BEA by practitioners in schools. One study surveyed a sample of school psychologists and found 70 % of school psychologists who responded received little or no training BEA procedures (Chafouleas et al. 2003). However, given a description of BEA, school psychologists rated BEA to be as acceptable for use in practice as norm referenced assessments, in which 95 % of the respondents indicated they had received training (Chafouleas et al. 2003). Thus, BEA use may increase with additional training.

Practical Questions

Despite the emergence and refinement of BEA technology for identifying effective individualized interventions for students struggling in an academic area, practical questions still remain. To begin, there is no unified approach to evaluating or selecting interventions. Although the variety of different approaches to BEA that appear in the literature suggest the flexibility of BEA, the lack of a cohesive set of measures, decision rules for what specific interventions or aspects of interventions to test, and criteria for identifying the most promising intervention poses challenges for training, implementation, and implementation fidelity. Second, the specific skills that should be targeted in a BEA for a particular academic area have yet to be fully demonstrated. Identifying a skill that is both sensitive to the effects of intervention during a brief analog analysis but also predictive of longer-term success in the learning environment requires careful consideration and empirical testing. Third, under what conditions is a BEA warranted? Despite the growing evidence that BEA is an effective method for identifying individualized interventions, it is unknown whether it should only be conducted after a supplemental intervention has been implemented unsuccessfully or whether there are some conditions (e.g., assessment results or data patterns) under which instructional staff should immediately conduct a BEA to identify an effective individualized intervention instead of first implementing a more standardized supplemental intervention.

BEA Within RTI

While the group design research literature is useful for identifying appropriate instruction and interventions that work on average for groups of students, especially in tiers 1 and 2, an idiographic approach is often needed for those who have not responded to standard treatment approaches. BEA is uniquely suited to meet this need as it matches student characteristics to instructional practices through experimental analysis, which could take the guesswork out of selecting an intervention because the initial analysis points practitioners in the right direction without extensive trial and error. Researchers suggest BEA for oral reading fluency takes approximately 60–90 min to complete (Burns and Wagner 2008), arguably far less time than a standard special education evaluation. While some would suggest that this level of individualization is the hallmark of special education, the authors argue that BEA could be used in tier 3 before a referral to special education is even made. Selecting and implementing appropriate interventions in tier 3 could prevent costly special education referrals and placements by making student success more likely. Finally, the second author's experience implementing BEA in local schools as part of a university supported research and service program found that undergraduate students with limited experience were successfully trained to conduct BEAs and to implement indicated interventions with high fidelity (i.e., 95 % on average). Thus, school personnel with expertise in some of the basic tenets of BEA could easily be trained to conduct them (e.g., school psychologists). Moreover, the use of BEAs to identify tier 3 interventions could happen within a team context of teams in which trained practitioners such as school psychologists would guide the use of the data.

Directions for Future Research

Although the evidence supporting the utility of BEA is growing, there are ample directions for future research. Relatively few studies have shown the extended effects of BEA-identified interventions in the areas of math, written language, or other academic skills besides reading. In addition, there is a broad range and many subtle nuances of interventions within and across academic skills that need to be more fully examined. For example, what other math interventions might be tested for different types of math problems and what different types of contingencies might be relevant in a BEA, and under what conditions might one be expected to be more effective than another? What skills should be measured during acquisition, fluency, generalization, and adaptation in various academic skill areas?

Another set of issues pertains to which member(s) of the instructional team in a school should be responsible for conducting a BEA with a student. Is it the responsibility of the general education teacher, the special education teacher, the school psychologist? Although it is rare for a BEA to be in any one staff person's job description, arguably the responsibility for prevention of school failure lies with all instructional and support staff. The issue of who is responsible for conducting the BEA is also related to issues pertaining to training and expertise. The concepts and procedures that are essential to BEA are rarely taught in preservice teacher preparation programs for general education teachers, and even more rarely for secondary general education teachers. However, if the student is not being referred for or currently does not receive special education services, it is likely that neither the school psychologist nor the special education teacher have the resources to complete a BEA. Thus, the question remains, which professional educator(s) in the school has or should have the training and expertise to conduct such individualized analyses in school settings?

Finally, also unclear are the differential outcomes for students who receive a BEA compared to those who receive standard treatment. As mentioned previously, unknown are the effectiveness on student outcomes and the efficiency of practices that involve supplemental interventions, standardized intensive interventions, or BEA-identified individualized interventions.

Conclusion

BEA is a useful tool for practitioners seeking to match interventions to student need in tier 3, especially for oral reading fluency. A wide variety of intervention approaches have been examined including antecedent interventions such as various forms of modeling, practice, choice, and adjusting the difficulty level of the material. Consequence-based interventions such as error correction and incentive for increased performance have also been used, as well as combinations of the two types of interventions. Limited research has also demonstrated that BEA can be applied to different academic skills including early literacy skills, math computation, letter formation, and written expression. Because BEA is a dynamic assessment tool guided by student response to intervention, it is potentially adaptable for use with a wide variety of independent and dependent variables, but more research is needed. Not much is known about the feasibility of using BEA in school contexts, but as more and more schools adopt multi-tiered systems of support, it is apparent that BEA is a viable tool for guiding decisions about what to teach for students who have not responded to standard approaches of skill remediation.

References

Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2,* 5–142.

Bonfiglio, C. M., Daly, E. J. III, Marten, B. K., Lin, L.-H., & Corsaut, S. (2004). An experiemtnal analysis of reading interventions: Generalizations acorss instructional strategies, time, and passages. *Journal of Applied Behavior Analysis, 37,* 111–114.

Burns, M. K., & Coolong-Chaffin, M. (2006). Response-to-intervention: Role for and effect on school psychology. *School Psychology Forum, 1,* 1–13.

Burns, M. K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37,* 126–136.

Burns, M. K., Ganuza, Z., & London, R. M. (2009). Brief experimental analysis of written letter formation: Single-case demonstration. *Journal of Behavioral Education, 18,* 20–34.

Chafouleas, S. M., Riley-Tillman, T. C., & Eckert, T. L. (2003). A comparison of school psychologists' acceptability, training, and use of norm-referenced, curriculum based, and brief experimental analysis methods to assess reading. *School Psychology Review, 32,* 272–281.

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *Communique, 34*(1) 6–7.

Codding, R., Baglici, S., Gottesman, D., Johnson, M., Kert, A., & Lebeouf, P. (2009). Selecting intervention strategies: Using brief experimental analysis for mathematics problems. *Journal of Applied School Psychology, 25,* 146–168.

Daly, E. J., III., & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27,* 459–469.

Daly, E. J., III., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance. *School Psychology Review, 26,* 554–574.

Daly, E. J., III., Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8,* 203–218.

Daly, E. J., III., Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 82–94.

Daly, E. J., III., Murdoch, A., Lillenstein, L., Webber, L., & Lentz, F. E. (2002). An examination of methods for testing treatments: Conducting brief experimental analyses of the effects of instructional components on oral reading fluency. *Education and Treatment of Children, 25,* 288–316.

Daly, E. J., III., Persampieri, M. J., McCurdy, M., & Gortmaker, V. J. (2005). Generating reading interventions through experimental analysis of academic skills: Demonstration and empirical evaluation. *School Psychology Review, 34,* 395–414.

Daly, E. J., III., Garbacz, S. A., Olson, S. C., Persampieri, M., & Ni, H. (2006). Improving oral reading fluency by influencing students' choice of instructional procedures: An experimental analysis with two students with behavioral disorders. *Behavioral Interventions, 21,* 31–60.

Daly, E. J., III., Bonfiglio, C. M., Mattson, T., Persampieri, M., & Foremann-Yates, K. (2006). Refining the experimental analysis of academic skills deficits: Part I. An investigation of variables that affect generalized oral reading performance. *Journal of Applied Behavior Analysis, 38,* 485–497.

Eckert, T. L., Ardoin, S. P., Daisey, D. M., & Scarola, M. D. (2000). Empirically evaluating the effectiveness of reading interventions: The use of brief experimental analysis and single-case designs. *Psychology in the Schools, 37,* 463–474.

Eckert, T. L., Ardoin, S. P., Daly, E. J., III., & Martens, B. K. (2003). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35,* 271–281.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-Intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18,* 157–171.

Gortmaker, V. J., Daly, E. J., III, McCurdy, M., Persampieri, M. J., & Hergenrader, M. (2007). Improving reading outcomes for children with learning disabilities: Using brief experimental analysis to develop parent-tutoring interventions. *Journal of Applied Behavior Analysis, 40,* 203–221.

Gray, B. B., Baker, R. D., & Stancyk, S. E. (1969). Performance determined instruction for training in remedial reading. *Journal of Applied Behavior Analysis, 2,* 255–263.

Hart, B. M., & Risley, T. R. (1968). Establishing use of descriptive adjectives in the spontaneous speech of disadvantaged preschool children. *Journal of Applied Behavior Analysis, 1,* 109–120.

Jones, K. M., & Wickstrom, K. F. (2002). Done in sixty seconds: Further analysis of the brief assessment model for academic problems. *School Psychology Review, 31,* 554–568.

Jones, K. M., Wickstrom, K. F., Noltemeyer, A. L., Brown, S. M., Schuka, J. R., & Therrien, W. (2009). An experimental analysis of reading fluency. *Journal of Behavioral Education, 18,* 35–55.

Kennedy, C. H. (2005). *Single-case designs for educational research*. New York: Allyn and Bacon.

Kratochwill, T. R., & Bergan, J. R. (1990). *Behavioral consultation in applied settings*. New York: Plenum Press.

Malloy, K. J., Gilbertson, D., & Maxfield, J. (2007). Use of brief experimental analysis for English language learners. *School Psychology Review, 36,* 291–310.

McComas, J. J. (1994). *Instructional strategy assessment: Structural analyses of academic performance.* Doctoral Dissertation, University of Iowa.

McComas, J. J., Wacker, D. P., & Cooper, L. J. (1996). Experimental analysis of academic performance in a classroom setting. *Journal of Behavioral Education, 6,* 121–130.

McComas, J. J., Wacker, D. P., Cooper, L. J., Asmus, J. M., Richman, D., & Stoner, B. (1996). Brief experimental analysis of stimulus prompts for accurate responding on academic tasks in an outpatient clinic. *Journal of Applied Behavior Analysis, 29,* 397–401.

McComas, J. J., Wagner, D., Coolong-Chaffin, M., Holton, E., McDonnell, M., & Monn, E. (2009). Prescriptive analysis: Further individualization of hypothesis testing in brief experimental analysis of reading fluency. *Journal of Behavioral Evaluation, 18,* 56–70.

National Center on Response to Intervention. (2013). http://www.rti4success.org. Accessed Nov 2013.

Noell, G. H., Gansle, K. A., Witt, J. C., Freeland, J. T., & LaFleur, L. H., et al. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659–663.

Noell, G. H., Freeland, J. T., Witt. J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39,* 335–355.

Northup, J., Wacker, D., Sasso, G., Steege, M., Cigrand, K., Cook, J., & DeRaad, A. (1991). A brief functional analysis of aggressive and alternative behavior in an outclinic setting. *Journal of Applied Behavior Analysis, 24,* 509–522.

Parker, D. C., Dickey, B. M., Burns, M. K., & McMaster, K. L. (2012). An application of brief experimental analysis with early writing. *Journal of Behavioral Education, 21,* 329–349.

Persampieri, M., Gortmaker, V., Daly, E. J., III, Sheridan, S. M, & McCurdy, M. (2006). Promoting parent use of empirically supported interventions: Two experimental investigations of child outcomes. *Behavioral Interventions, 21,* 31–57.

Petursdottir, A.-L., McMaster, K., McComas, J. J., Bradfield, T., Braganza, V., Koch-McDonald, J., Rodriguez, R., & Scharf, H. (2009). Brief experimental analysis of early reading interventions. *Journal of School Psychology, 47,* 215–243.

Tilly, W., & David, III. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes *Best practices in school psychology V*. Bethesda: The National Association of School Psychologists.

VanAuken, T., Chafouleas, S. M., Bradley, T. A., & Martens, B. K. (2002). Using brief experimental analysis to select oral reading interventions: An investigation of treatment utility. *Journal of Behavioral Education, 11,* 165–181.

VanDerHeyden A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education, 18,* 71–91.

Wagner, D., McComas, J. J., Bollman, K., & Holton, E. (2006). The use of functional analysis of academic response to intervention for oral reading. *Assessment for Effective Intervention, 32,* 40–49.

Whitlock, C. (1966). Note on reading acquisition: An extension of laboratory principles. *Journal of Experimental Child Psychology, 3,* 83–85.

Wilbur, A., & Cushman, T. P. (2006). Selecting effective academic interventions: An example using brief experimental analysis for oral reading. *Psychology in the Schools, 43,* 79–84.

Tier 3: Intensive Mathematics Intervention Strategies

Robin S. Codding and Ryan Martin

Mathematics has often been eclipsed by literacy in academic intervention research, despite the importance of both skills for a child's overall success (Methe et al. 2011). Knowledge of early mathematics concepts was one of the most powerful predictors of future academic achievement (Duncan et al. 2007). More broadly, mathematics proficiency is necessary for individuals to successfully navigate the modern world. Principles of mathematics are embedded in simple daily tasks such as cooking or a trip to the grocery store, and also in more advanced, yet equally important, tasks like paying taxes or balancing a household budget. Competence in mathematics is fundamental in science and engineering careers, which are predicted to outpace general job growth in the coming years (National Mathematics Advisory Panel 2008). Additionally, those with poor skills in mathematics are less likely to be involved within their communities, have difficulty managing finances, and experience various employment challenges (Methe et al. 2011). It is critical that schools address mathematics intervention through evidence-based methods to support positive student outcomes in such a foundational skill.

This chapter reviews previous research that identifies types of mathematics challenges experienced by students, and provides a general overview of the status of mathematics interventions that focus on whole number knowledge. Specific discussion of interventions that address rational numbers or applied problem-solving extends beyond the scope of this chapter. Second, the chapter describes core features of tier 3 intervention packages that can be utilized across mathematics topics, while highlighting their application to number combinations and computation fluency using examples from our own intervention work. Finally, the chapter discusses future areas of needed research.

R. S. Codding (✉) · R. Martin
University of Massachusetts-Boston, Boston, USA
e-mail: Robin.Codding@umb.edu

Review of Previous Research

Mathematics is a complex content area consisting of many topics that are reflected in the challenges experienced by students. As many as 30 different behaviors differentiated children with mathematics difficulties from typically performing peers (Bryant et al. 2000). Thus, let us begin by discussing student mathematics difficulties and provide an overview of mathematics interventions.

Description of Student Mathematics Difficulties

Solving word problems represents a commonly cited area of difficulty, and some of the best predictors of mathematics challenges include procedural difficulty with multi-step problems and borrowing errors (Bryant et al. 2000; Geary

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_22

1993). Children with mathematics learning disabilities also tend to use less mature counting strategies, such as guessing and the sum procedure (e.g., in the problem 3 + 4, children line up two sets of items and count all objects starting from 1) and make more counting errors than their peers (Geary et al. 2004). This may result because children are not familiar with the magnitude sequence of numerals and can only determine which number comes first (e.g., 3 or 4) by counting from one (Baroody et al. 2009).

A series of longitudinal studies illustrated that knowledge of and fluency with number words, numerals and their quantities, as well as number lines represent critical components that underlie mathematics learning (Geary 2011; Geary et al. 2009; Jordan et al. 2009). Difficulties in these core areas have been shown to be stable through the elementary school years, and mathematics competency in kindergarten is highly predictive of later school achievement (Duncan et al. 2007; Jordan et al. 2009; Morgan et al. 2009).

A robust finding in the literature is that challenges with automatic fact retrieval and calculation fluency may be the unifying feature of children who struggle in mathematics (e.g., Geary 2011; Gersten et al. 2005). Research has illustrated that students without mathematics learning disabilities recall as many as three times the number of basic facts as their peers with learning disabilities even though accuracy of basic fact performance is equivalent (Hasselbring et al. 1988; Gersten and Chard 1999), which is consistent with research that found that poor performance with simple arithmetic tasks at the end of kindergarten predicts learning disabilities by the end of third grade (Mazzocco and Thompson 2005).

Students without basic fact fluency may be less able to grasp underlying mathematics concepts or access higher-level mathematics curricula (Gersten and Chard 1999). Students that have to allocate more cognitive resources to retrieving the solution to 9 + 6 (for example) may experience interference with higher-order thinking or problem-solving because directing conscious attention towards multiple tasks simultaneously is challenging (Barrouillet and Fayol 1998; Dahaene 1997). Evidence also exists to suggest that students with low number combination fluency may exhibit greater anxiety for mathematics tasks than students with more fluent skills (Cates and Rhymer 2003). Whether it is the additional effort required, anxiety, or a combination of the two, collectively this evidence suggests that students without number combination fluency may engage in less frequent practice with mathematics content and complete fewer mathematics-related tasks (Billington et al. 2004; Skinner et al. 1997).

Overview of Mathematics Interventions

With the recognition that computational fluency is a key aspect of mathematics learning, as well as research supporting that its absence represents a hallmark feature of mathematics learning disabilities (Gersten et al. 2005), the NMAP report (2008) indicated that automaticity with number combinations is a necessary goal for all children. Mathematics computation and applications have been shown to be distinct albeit highly related ($r = 0.83$) constructs suggesting that skills in one area are necessary for success in the other (Thurber et al. 2002). Jordan et al. (2003) illustrated that students with low mastery of mathematics facts make minimal, if any, progress on a variety of other tasks compared to higher-mastery groups suggesting the critical role that computation fluency has on building procedural and conceptual knowledge of mathematics principles. More intervention research has been conducted on this aspect of mathematics than with rational numbers, early number knowledge, or word problem-solving.

Preliminary intervention research targeting early numeracy has focused on developing individualized and small-group interventions that include instruction on strategic counting (e.g., counting up; doubles + 1) and magnitude representations through practice with magnitude comparisons (e.g., which number is bigger?), number games, and number-line estimation tasks (Bryant et al. 2011; Fuchs et al. 2005; Seigler and Ramani 2009). Research examining word problem-solving interventions has produced convincing evidence implicating the use of schema-based instruction, which directly teaches underlying structures of word problems through problem mapping and diagramming (Fuchs et al. 2007; Jitendra 2007). These data suggest that there are promising protocols that interventionists can incorporate within their school systems; however, more research on intervention effectiveness is needed.

In a review of the literature on mathematics computation intervention research between 1980 and 2007, the 12 different strategies listed in Table 1 were identified, 50 % of which were analyzed in more than one study (Codding et al. 2009b). These interventions were categorized as either simple intensity interventions that improve the academic learning environment, or moderate intensity reflecting a specific alteration in the form or type of instruction provided (Barnett et al. 2004; Shapiro 2011). All of these interventions are resource friendly requiring minimal training and materials. Generally, these interventions were effective, but for some improvement depended on the outcome measure (i.e., percentage accuracy, fluency (digits correct), or general

Table 1 Selected mathematics intervention strategies and descriptions

Intervention strategy	Intervention description	Effect size ranges across outcome type		
		% Correct	D/PCM	Score
Reinforcement and performance feedback				
Earning contingent free time	Free time is earned contingent upon a predetermined goal or completing a specified number of problems	4.71		
Goal setting with and without reinforcement	Self- or teacher-set goals are established, and student is provided with data illustrating his/her progress toward goals. Positive reinforcement can be provided incrementally or upon reaching goals		0.00–1.20	0.65–2.67
Flash card drill				
Traditional drill	A sequence of math facts is presented to the student; the student provides responses orally and receives corrective feedback when appropriate		1.12–2.58	
Incremental rehearsal	Unknown math facts are identified and presented to the student one at a time. The student repeats the fact with the answer, and receives corrective feedback when appropriate. Known facts are then folded in according to a 9:1 ratio and rehearsed with unknown facts		3.42–17.00	
Self-management				
Self-instruction	A student is taught various self-questioning or "think-aloud" strategies and/or simple heuristics to solve problems independently. Strategies are often modeled and prompted by the teacher	0.84–19.50	0.31–5.53	0.90
Self-monitoring	The student is taught to use checklists or other forms of self-evaluation at set intervals to monitor his/her behavior according to a target behavior	−0.06–5.24		
Cover–copy–compare	The student looks at a math problem with the answer, then covers the problem and answer. The student records the problem and answer, then uncovers the original problem and compares the answers for accuracy	0.75–6.29	−0.34–7.72	

Table 1 (continued)

Intervention strategy	Intervention description	Effect size ranges across outcome type		
Cue cards and graphing	Cue cards contain steps for each part of solving a problem. The teacher models cue cards for solving problems, then students use them to solve problems. The student graphs his/her performance each session	3.01–6.32		
Peer tutoring	Various procedures in which students are taught specific techniques for acting as instructional partners. Students work together to learn and practice skills		0.17–5.98	0.17–0.30
Taped problems	A tape-recording of math problems is created to correspond to problems on a worksheet. Answers to the problems are provided on the recording after a brief time delay. The student completes the worksheet as the recording is played, and writes his/her answer during the time delay. If the student fails to answer or provides an incorrect response, he/she writes in the correct answer upon hearing it	1.52–4.41	0.87–4.41	
Count by's (multiplication)	The student is taught to solve multiplication problems by skip counting by the multiplicand (i.e., solving 3 × 4 by counting 3, 6, 9, 12). The student repeats the counting scheme until he/she can do so at a rate of one count per second		3.30–4.74	
Explicit timing	The student completes as many problems as possible in 1 min. At the end of the interval, the student is instructed to stop, underline the last number written, and put his/her pencil in the air. This procedure continues throughout a designated practice time (e.g., 5 min)	0.01–0.08	0.20–1.14	

These strategies and their descriptions resulted from review of the literature according to Codding et al. (2009b) and do not represent a comprehensive list of all possible mathematics intervention strategies. *D/PCM* digits or problems correct per minute. Score was yielded from criterion, norm-referenced, or researcher-generated tests

mathematics achievement (according to criterion- or norm-referenced tests)). These findings suggest that replication of effectiveness findings is warranted and that knowing what intervention strategy to implement is necessary but not sufficient. Knowledge of student-specific skill levels and appropriate treatment match according to level of skill proficiency might also be useful, particularly for students experiencing the most difficulty.

Given the paucity of standard protocol mathematics interventions (Fuchs et al. 2008; National Mathematics Advisory Panel 2008), this chapter will emphasize common intervention features that have been broadly supported in the literature and can be used to generate individualized treatment plans for students in need of tier 3 supports. Examples are used from our research on whole-number computation fluency and instructional decision-making to illustrate conceptualization and application of relevant treatment components.

Key Features of Tier 3 Interventions

Common intervention features identified through the extant literature offer promising guidelines for generating treatment packages for use with students exhibiting some of the greatest problems in mathematics (Burns et al. 2010; Codding et al. 2011a; Fuchs et al. 2008; Baker et al. 2002; Gersten et al. 2009; Swanson 2009; Swanson and Sachse-Lee 2000). These core components include: (a) matching treatment to skill needs, (b) explicit instruction, (c) self-instruction, (d) concrete–representational–abstract (C–R–A) instructional sequencing, (e) providing productive opportunities to practice, and (f) incorporating motivation (Table 2). Independently, these

Table 2 Key components of tier 3 mathematics interventions

Components	Description
A. Treatment by skill match	Two-step process for matching treatment with students' level of skill proficiency
Curriculum-based assessment—instructional design	Used to identify students' specific skill strengths and weaknesses
Instructional hierarchy	Decision-making heuristic to match skill levels with intervention strategies
B. Explicit instruction	Method of teaching that subdivides skills into smaller, more manageable steps in a planned sequence with the incorporation of many examples
C. Self-instruction	Method of teaching that requires students to monitor their own problem-solving through the use of teacher-modeled visual and verbal prompts
D. Concrete, representation, and abstract sequenced instruction	Method of teaching that incorporates concrete manipulatives, visual representations, and concrete numerals
E. Opportunities to practice	
Drill	Rehearsal of isolated items
Practice with modeling	Use of newly learned responses in a different context or combined with previously learned responses where the correct response is modeled
Cumulative review	Rehearsal of learned responses or sequenced review of previously learned skills
F. Motivation	
Direct	Contingent on mathematics performance (e.g., digits correct)
Indirect	Contingent on task initiation, persistence, and/or completion

intervention components seldom demonstrate efficacy; however, combinations of these strategies have consistently led to positive student outcomes among children struggling with mathematics (Codding et al. 2011a; Fuchs et al. 2008; Gersten et al. 2009; Swanson 2009).

Treatment by Skill Match

Additional assessment and analysis is recommended when developing tier 3 supports, given previous challenges these students have experienced in core instruction and potential lack of response to tier 2 interventions. Rather than simply identifying available interventions, it might be useful to apply a problem-solving approach so that treatment selection matches student level of skill development. Selecting treatments according to their match with students' level of skill proficiency requires two steps: (a) identify students' specific skill strengths and weaknesses and (b) use a decision-making heuristic to match skill level with an intervention strategy (Burns et al. 2010; Codding et al. 2007; Daly et al. 2000). Tier 3 interventions should use the hierarchical nature of mathematics content to emphasize prerequisite skills that may not necessarily match grade-level content (Gersten et al. 2009; Swanson 2009). For example, fourth-grade students with unmastered addition and subtraction skills should receive intervention support that addresses these skill weaknesses prior to focus on multiplication or division (Shapiro 2011). Within a tiered system of service delivery, all students will have access to grade-level content during core instruction and the provision of tiered services emphasizing key foundational skills will facilitate access to this content.

Curriculum-based assessment (CBA) is a useful tool for intervention planning (Burns 2004). CBA-instructional design (Gickling and Havertape 1981) requires a survey-level assessment be constructed with a range of single-skill mathematics facts reflective of Common Core State Standards (National Governers Associaton Center for Best Practices, Council of Cheif State School Officers 2010) and/or National Council of Teachers of Mathematics (2006) focal point scope and sequence recommendations. CBA worksheets are administered beginning with grade-level skills and continue down the skill hierarchy with the goal of identifying skills that fall in the mastery, frustration, and instructional ranges of proficiency according to specified criteria. For example, Burns et al. (2006) derived criteria experimentally which resulted in scores below 14 (grades 2 and 3) and 24 (grades 4 and 5) digits

Table 3 Treatment by skill interaction: Combining CBA-ID and the instructional hierarchy for treatment selection. (CBA-ID criteria according to Burns et al. 2006)

CBA-ID criterion level	CBA-ID score	Stage of instructional hierarchy	Intervention characteristics	Intervention Strategies
Frustration	Grades 2 and 3	Acquisition	Concrete–representation–abstract Demonstration Modeling Guided practice Practice w/frequent, immediate and corrective feedback Prompting think aloud	Cover–copy–compare Explicit instruction Incremental rehearsal Traditional drill Self-instruction
	<14 DCPM[a]			
	Grades 4+			
	<24 DCPM[a]			
Instructional	Grades 2 and 3	Fluency	Frequent and novel practice opportunities Independent practice in context Peer practice Performance feedback w/ or w/o Goal setting Contingent reinforcement Timed and repeated practice in context	Computer-assisted instruction Cue cards Explicit timing Peer tutoring Self-monitoring Taped problems
	14–31 DCPM[a]			
	Grades 4+			
	24–49 DCPM[a]			

CBA-ID curriculum-based assessment instructional design, *DCPM* digits correct per minute
[a]Burns, M. K., VanDerHeyden, A. M., & Jiban, C. L. (2006). Assessing the Instructional Level for Mathematics: A Comparison of Methods. *School Psychology Review, 35,* 401–418.

correct per minute (DCPM) representing the frustration level, scores between 14 and 31 DCPM (grades 2 to 3) and 24 to 49 (grades 4 and 5) DCPM representing the instructional level, and mastery performance exceeding the highest score of those ranges (see Table 3). The skill selected for treatment focus will depend on the number of skills identified in the frustration and instructional ranges with a potential starting point being the unmastered skill lowest in the skill hierarchy.

Although there are a number of decision-making heuristics that could be selected to make an appropriate skill by treatment match (Christ 2008), our research has focused on the application of the instructional hierarchy (IH) as defined by Haring and Eaton (1978) and adapted to mathematics by Rivera and Bryant (1992). The IH suggests that skill development consists of four stages: acquisition, fluency, generalization, and adaption. Initially, students focus on acquiring new skills by developing accuracy. Therefore, performance of students in this stage is slow and inaccurate. Once students have achieved accuracy, the next step is to build fast and accurate performance with the skill. Haring and Eaton (1978) postulated that accurate and fluent skill performance permits generalization of skill use over time, under different conditions, and with new stimuli (the third stage of skill development), and, lastly, skill proficiency results in adaption of skills to novel mathematics tasks. When used with CBA-ID, performance that falls in the frustration range may represent the acquisition stage of skill development and performance in the instructional range may represent the fluency stage (Burns et al. 2010).

Intervention techniques for students in the acquisition stage of skill development might include modeling, guided practice, think-alouds, and immediate corrective feedback using concrete manipulatives, visual representation, and numerals (Rivera and Bryant 1992). However, once students have attained accuracy, continued use of these strategies may be counterproductive for building fluency, as students may become strategy dependent (Poncy et al. 2006). Characteristics of interventions for students in the fluency stage of skill development include novel and frequent opportunities to practice recently acquired skills through the use of timed drills, peer tutoring, reinforcement, goal setting, reinforcement, and computers.

The manner in which skill generalization can be facilitated is less clear. In fact, 38 % of studies included in recent meta-analytic research-measured skill retention and only 11 % measured generalization (Codding et al. 2009b). There is some evidence that suggests high rates of skill fluency result in subsequent proficiency on nationally normed achievement tests (VanDerHeyden and Burns 2008) and strong correspondence has been yielded between single and multiple-digit multiplication fluency (Lin and Kubina 2005). We have also found that higher rates of fluency in single-skill subtraction resulted in generalization with 2 × 1 digit subtraction with regrouping problems as compared to students displaying lower levels of fluency (Codding et al. 2009a) and that improved performance on multiplication flash cards using a drill procedure resulted in improved fluency on multiplication and fraction curriculum-based measurement (CBM) probes (Codding et al. 2010). These data suggest that adequate levels of fact fluency on key component skills might facilitate problem-solving on composite skills (Johnson and Layng 1992; Skinner and Daly 2010). It may also be that arranging intervention opportunities to facilitate generalization such as through the use of multiple examples or training hierarchical skill elements simultaneously promotes generalized learning (Codding and Poncy 2010; Skinner and Daly 2010). Applied practice activities within learning centers or instructional games could also prove effective (Rivera and Bryant 1992).

Two studies were conducted that offer some preliminary support for using a skill by treatment (CBA−ID + IH) approach to intervention selection. In our initial study (Codding et al. 2007), second- and third-grade students were randomly assigned to one of three groups: Explicit Timing, Cover–Copy–Compare (CCC), or Control. Explicit timing was hypothesized to be an intervention that would benefit students in the fluency stage of the IH, and CCC was selected to represent the acquisition phase of the IH. The length of practice time was held constant and occurred twice weekly for 6 weeks. Using multilevel modeling, an interaction between initial CBA score and treatment was found whereby students whose performance fell in the instruction range (i.e., fluency stage of IH) prior to inception of the study and received explicit timing made more growth during the project and achieved higher final performance than students with comparable performance receiving either no treatment or CCC. Students whose initial scores fell in the frustration range (i.e., acquisition stage of the IH) and received explicit timing had lower scores at the end of treatment and flatter rates of growth than comparable students also in the acquisition stage of the IH assigned to the control condition or receiving the CCC treatment. This supported the notion that students performing in the instruction range of the skill selected for treatment (i.e., fluency stage of IH) benefited from a treatment procedure that contained strategies consistent with fluency building but this intervention was not effective when applied to students whose skills were in the frustration range (i.e., acquisition stage of IH).

Next, (Burns et al. 2010) a meta-analysis was conducted using single-case design studies that employed mathematics computation interventions. Intervention strategies were categorized as consistent with building accuracy (i.e., acquisition stage of IH) or fluency (i.e., fluency stage of IH) using criteria from Rivera and Bryant (1992). Baseline scores were analyzed to determine whether each participant's CBA score fell in the frustration or instructional range. We were interested in determining whether the indicated treatment (e.g., acquisition strategies applied when CBA score fell in frustration range) would result in better performance than the contraindicated treatment (e.g., acquisition strategies applied when CBA skill level was instructional). The obtained studies included 55 participants in grades 2–6 with 65.5 % displaying CBA scores in the frustration range and 34.5 % displaying CBA scores in the instructional range. We were unable to evaluate the effect of fluency-building interventions on students whose performance was in the instructional range of CBA (indicated treatment) due to a limited number of participants that fell in this category. Our hypotheses were partially supported (see Fig. 1). It was found that acquisition building strategies applied to students with CBA scores in the frustration range resulted in large effect sizes (indicated treatment) whereas fluency building strategies applied to

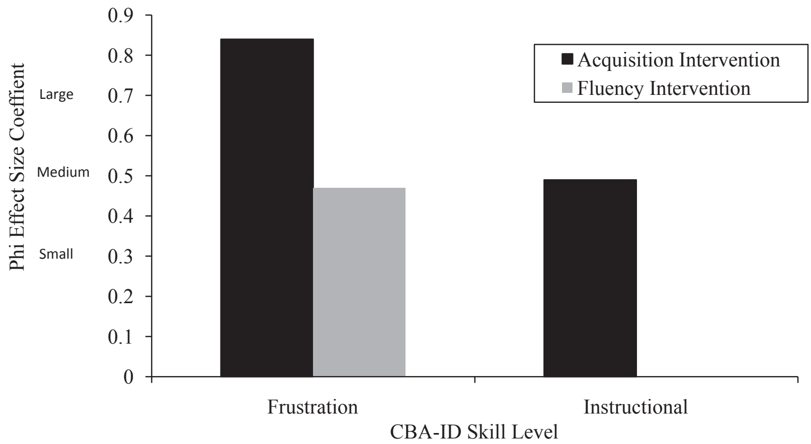


Fig. 1 Mean phi coefficient for skills of students falling in the frustration and instructional ranges and receiving intervention strategies consistent with either acquisition or fluency stages of the instructional hierarchy. (Burns et al. 2010)

students with similar performance resulted in small to moderate effect sizes (contraindicated treatment). For students with CBA scores in the instruction range receiving an acquisition intervention a small effect size was yielded (contraindicated treatment).

Collectively, these data suggest that using CBA-ID with the IH may be a useful guide to select an appropriate treatment match for students. That being said additional research is warranted in this area. It is unclear whether skill development occurs in the discrete stages suggested by the IH (Martens and Eckert 2007). It also remains to be seen whether treatment strategies should be applied sequentially so that when a skill is in the acquisition stage one set of techniques is provided but once achieved another set of techniques is employed for fluency building, or, whether treatment packages should be developed with both sets of techniques applied simultaneously (Skinner and Daly 2010).

Explicit Instruction

In no uncertain terms, explicit instruction has been described as a critical instructional method for helping children acquire basic skills in mathematics, particularly when providing tier 3 services (Baker et al. 2002; Fuchs et al. 2008; Gersten et al. 2009; Swanson 2009; Swanson and Sachse-Lee 2000). Explicit instruction includes the teaching of specific skills and strategies by breaking them down into smaller, more manageable steps in a planned sequence with the incorporation of many examples. Typically, this process begins with modeling, error correction, and guided practice and gradually shifts to independent practice using teacher modeling and corrective feedback. This intervention feature provides an effective foundation for helping students understand concepts that underlie number combinations such as quantity discrimination, counting, and mental use of number lines (Gersten et al. 2005; Poncy et al. 2006).

Self-Instruction

Self-instruction, also considered a self-regulation or meta-cognitive strategy, is intended to teach students to monitor their own problem-solving through the use of teacher-modeled visual and verbal prompts (Goldman 1989). These interventions are often implemented by students themselves or in conjunction with support from a teacher and have demonstrated large effect sizes, as opposed to treatments implemented solely by a teacher (Codding et al. 2011a). Specific self-instruction strategies for students often consist of "think-alouds," self-questioning, and simple heuristics; for example, the "Say–Ask–Check" method encourages students to read a problem, ask themselves questions about the problem, and check their work at each step (Kroesberg and Van Luit 2003; Montague 2008). Similarly, steps for solving different types of problems can be stated aloud by students in action terms, such as “read the problem first, next find the bigger number,” followed by students repeating and completing successive steps (Kroeger and Kouche 2006; Tournaki 2003). Accurate verbalization of problem-solving steps is particularly important for struggling students (Gersten et al. 2009; Siegler and Booth 2004; Swanson 2009), as such strategies help students to focus attention to the task, prevent impulsive responding, and manage frustration (Meichenbaum and Goodman 1971).

C–R–A Sequence

Mathematics learning opportunities that incorporate the use of visual representations and concrete manipulatives along with numerals is an important feature of effective interventions with struggling students (Flores 2010; Gersten et al. 2009; Miller and Mercer 1993; Mercer and Miller 1992; Swanson 2009). The use of manipulatives such as counters, chips, and blocks is common in mathematics instruction; however, current recommendations suggest that manipulatives are most effective when used sequentially with concrete examples, visual representations, and abstract numerical symbols (Baroody et al. 2009). Mercer and Miller (1992) demonstrated that students with mathematics learning disabilities benefited from using concrete manipulatives an average of three times before practicing the same concept using visual displays such as pictures of objects, number lines, tally marks, and/or ten frames. It is essential that the use of visuals and manipulatives be implemented according to a progression, such that manipulatives are used first and faded to proceed to the use of visual representations, which serve as an intermediate step. When students are able to accurately and independently solve problems with visual representations, more advanced practice with abstract symbols (i.e., numerals) should be initiated (Flores 2010).

Productive Opportunities to Practice

US students cannot solve basic facts as quickly or efficiently as their international peers which is due, in part, to the quantity and quality of practice provided within the classroom and offered via traditional mathematics textbook curricula (National Mathematics Advisory Panel 2008). Educators tend to focus on accurate mathematics performance as opposed to rates of responding and tend to emphasize complex composite mathematics skills at the expense of component prerequisite skills (Daly et al. 2007). This has resulted in rapid introduction of new mathematics concepts without supplying sufficient opportunities to master content and ensure retention of previously learned skills (Daly et al. 2007; National Mathematics Advisory Panel 2008). However, development and implementation of carefully constructed direct practice opportunities is widely supported in the literature (Codding et al. 2011a; Daly et al. 2007; Powell et al. 2009; Swanson 2009). Practice has been described as a necessary treatment component for students receiving tier 3 services (Binder 1996; Fuchs et al. 2008) and treatment packages that contain direct opportunities for practice produce better outcomes that those without (Powell et al. 2009). Binder (1996) suggested that fluency building might be facilitated by allocating as much as 70 % of instructional time to practice activities.

Practice consists of a number of activities including drill, practice, and cumulative review. Drill has been defined by Haring and Eaton (1978) as the rehearsal of isolated items such as with flash cards, whereas practice requires the use of newly learned responses in a different context (i.e., worksheets, word problems; Daly et al. 2007; Fuchs et al. 2008) or combined with previously learned responses as might be seen in a mathematics worksheet. Cumulative review can be conceptualized as rehearsal of learned responses, such as practicing all one by one digit addition number combinations (Fuchs et al. 2008) or sequenced review of skills previously learned to a criterion until all skills in a hierarchy or components of a more complex skill are mastered (Engelmann and Carnine

1982). In our (Codding et al. 2011a) meta-analysis of basic fact fluency interventions, the impact of various practice activities on student outcomes were coded and isolated. Our findings demonstrated that for students in need of intensive interventions, practice that incorporated modeling of the problem with the correct answer or a flash-card drill procedure where the ratio of known to unknown facts was controlled resulted in median gains of 9.75 DCPM and 25.33 DCPM, respectively, compared to median gains of 2 DCPM for practice without modeling. However, this does not necessarily mean that only one type of practice activity should be used. It has been postulated that drill facilitates skill fluency, whereas as other practice activities lead to retention, maintenance, and generalization (Daly et al. 2007; Haring and Eaton 1978) and our study only evaluated immediate effects of treatment.

Development of effective practice opportunities requires that students be presented with material that matches their instructional level. In mathematics, this means that students should perform skills with approximately 70–85 % accuracy (Gickling and Thompson 1985) and/or exhibit rates of performance that range from 14 to 31 DCPM (grades 2 and 3), or 24 to 49 DCPM (grades 4 and 5; Burns et al. 2006). Practice sessions should be organized to be brief and frequent lasting approximately 10 min (Daly et al. 2007; Gersten et al. 2009). Practice should be sequenced systematically in small sets and according to student progress (Fuchs et al. 2008; Martens and Eckert 2007; Woodward 2006). There is some evidence to suggest that constructing small practice sets according to fact families (i.e., $2 \times 3 = 6$; $6/2 = 3$) or introducing facts sequentially is beneficial (i.e., 0, 1, 2s, etc.; Hasselbring et al. 2006).

Motivation

The NMAP report (2008) highlighted the importance of incorporating strategies that facilitate motivation within treatment packages in order to encourage engagement in mathematics tasks. As students become accurate in mathematics, reinforcement for continued practice (as is recommended to build fluency) in the natural environment may diminish (Skinner et al. 1997). Students may in turn be less motivated to engage in overlearning, and rates of responding may not improve to levels necessary for skill retention and generalization (Daly et al. 2007). These problems are exacerbated for students in need of more intensive interventions (Fuchs et al. 2008) who are at increased risk of experiencing mathematics anxiety (Cates and Rhymer 2003) and less likely to persist on mathematics tasks given their difficult educational history (Billington et al. 2004). Encouragement provided to students engaging in mathematics tasks can be directly associated with mathematics performance (e.g., meeting a specified criterion, beating a score) or indirect, focusing instead on self-regulation with mathematical tasks (e.g., task persistence, initiation, or completion). The can’t do/won’t do assessment (VanDerHeyden and Witt 2008) can distinguish among students displaying skill deficits, performance deficits, and combined skill and performance deficits.

Token economies with and without response cost, where points or tokens are earned (or removed) and exchanged for prizes, can be incorporated into a treatment package. For example, *Math FLASH* (Fuchs et al. 2003), a computer-assisted program focusing on fact families, incorporates reinforcement by awarding students with stars for participating, attending, and providing effort toward each element of the intervention package. These stars are recorded on a chart after each treatment session and when 16 stars are accumulated, an assortment of prizes is presented from which the student selects one (Fuchs et al. 2008). Prizes could also be provided directly for improvement on performance such as digits or problems correct.

Providing students with feedback on their performance also improves student outcomes (Gersten et al. 2009). Performance feedback can consist of informing students of the number of correctly completed or answered problems, graphs depicting mathematics scores, or number and type of mastered skills. When feedback is linked to praise for effort, the effects are particularly powerful (Gersten et al. 2009). A series of studies have been conducted examining the added value of performance feedback with goal setting (PFGS) applied to treatment packages for kindergarteners (Codding et al. 2011b), third (Codding et al. 2009a), and fourth graders (Codding et al. 2005). For kindergarteners, a vertical bar chart was created with the scores indicated on the left side of the chart. Students were instructed to color in the bar chart after each session. In effect, the chart served as a thermometer with students getting closer to the top of the bar with each session of improved performance. A variety of crayons were used so that students could visualize not only the overall increased progress but also gains made each session.

For third and fourth graders, written and graphic performance feedback was provided. A PowerPoint generated graphic display of student progress using a bar chart with sessions on the *x*-axis and digits or problems correct on the *y*-axis were constructed. A praise statement was provided at the top of the graph when the student beat their score (e.g., "Way to go, you completed 2 more problems correct!"). A star was inserted on the graph at the point where the new performance goal (*y*-axis) met with the session number (*x*-axis). A text box below the star provided a goal statement ("Let's see if you can complete 30 problems correctly today!"). Students were shown the bar graph prior to each session, informed of their current performance levels, and provided with a performance goal that was visually depicted. In all cases, treatment packages with PFGS (i.e., Kindergarten Peer-Assisted Learning Strategies—KPALS + PFGS; Cover–Copy–Compare–CCC + PFGS) yielded significantly greater growth and higher final scores by study termination than treatment packages without PFGS (i.e., KPALS; CCC).

Summary and Areas for Future Research

Current research efforts are underway to develop new standard protocol mathematics interventions and refine existing ones to meet evidence-based expectations. Evaluation of general and specific intervention strategies that can improve whole and rational number knowledge, number combination fluency, and word problem-solving are needed. Fortunately, individualized treatment packages for students in need of tier 3 supports in mathematics can be generated through a problem-solving approach and developed according to the core treatment elements described in this chapter. These treatment components are resource friendly, offering options that require minimal additional training or materials for implementation.

Questions remain regarding the specific combination of treatment elements that are most effective including whether strategies to promote acquisition and fluency are introduced simultaneously or whether treatment strategies are altered depending on skill development. The extent to which existing strategies promote generalization, which is the ultimate goal of all tiered supports, or the manner in which to program for generalization when building treatment packages is unclear. Additional research is needed on the importance of skill sequences (easy to hard; component to composite) and how these arrangements might alter intervention efficiency. The interplay between outcomes (final performance, growth, trials to criterion) and treatment dose defined as number of treatment sessions per week, session duration, and total length of treatment is imperative for determining appropriate and efficient treatment options.

References

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103,* 51–73.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, R. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *The Journal of Special Education, 38,* 66–79.

Baroody, A. J., Bajwa, N., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews, 15*(1), 69–79. doi:10.1002/ddrr.45.

Barrouillet, P., & Fayol, M. (1998). From algorithmic computing to direct retrieval: Evidence from number and alphabetic arithmetic in children and adults. *Memory and Cognition, 26,* 355–368.

Billington, E. J., Skinner, C. H., & Cruchon, N. M. (2004). Improving sixth-grade students perceptions of high-effort assignments by assigning more work: Interaction of additive interspersal and assignment effort on assignment choice. *Journal of School Psychology, 42*(6), 477–490. doi:10.1016/j.jsp.2004.08.003.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *Behavior Analyst, 19,* 163–197.

Bryant, D., Bryant, B. R., & Hammill, D. D. (2000). Characteristic behaviors of students with LD who have teacher-identified math weaknesses. *Journal of Learning Disabilities, 33*(2), 168–177. doi:10.1177/002221940003300205.

Bryant, D., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78*(1), 7–23.

Burns, M. (2004). Using curriculum-based assessment in consultation: A review of three levels of research. *Journal of Educational and Psychological Consultation, 15,* 63–78.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35,* 401–418.

Burns, M., Codding, R. S., Boice, C., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instruction and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review, 39,* 69–83.

Cates, G. L., & Rhymer, K. N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education, 12*(1), 23–34. doi:10.1023/A:1022318321416.

Christ, T. J. (2008). Best practices in problem analysis. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 159–176). Bethesda: National Association of School Psychologists.

Codding, R. S., & Poncy, B. C. (2010). Introduction to the special issue: Toward an explicit technology for generalizing academic behavior. *Journal of Behavioral Education, 19,* 1–6.

Codding, R. S., Lewandowski, L. J., & Eckert, T. L. (2005). Examining the efficacy of performance feedback and goal setting interventions: A comparison of two methods of goal setting. *Journal of Evidence-Based Practices for Schools, 6,* 42–58.

Codding, R. S., Shiyko, M., Russo, M., Birch, S., Fanning, E., & Jaspen, D. (2007). Comparing mathematics interventions: Does initial level of fluency predict intervention effectiveness? *Journal of School Psychology, 45,* 603–617.

Codding, R. S., Chan-Iannetta, L., Palmer, M., & Lukito, G. (2009a). Examining a class-wide application of cover-copy-compare with and without goal setting to enhance mathematics fluency. *School Psychology Quarterly, 24,* 173–185.

Codding, R. S., Hilt-Panahon, A., Panahon, C., & Benson, J. (2009b). Addressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education and Treatment of Children, 32,* 279–312.

Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioral Education, 19,* 93–105.

Codding, R. S., Burns, M. K., & Lukito, G. (2011a). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research and Practice, 26*(1), 36–47. doi:10.1111/j.1540-5826.2010.00323.x.

Codding, R. S., Chan-Iannetta, L., George, S., Ferreira, K., & Volpe, R. (2011b). Early number skills: Examining the effects of class-wide interventions on kindergarten performance. *School Psychology Quarterly, 26,* 85–96.

Dahaene, S. (1997). *The number sense*: *How the mind creates mathematics*. New York: Oxford Univeristy Press.

Daly, E. J., Hintze, J., & Hamler, K. R. (2000). Improving practice by taking steps toward technological improvements in academic intervention in the new millennium. *Psychology in the Schools, 37,* 61–72.

Daly, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement instruction and intensity. *School Psychology Review, 36,* 562–581.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Engelmann, S., & Carnine, D. W. (1982). *Theory of instruction*: *Principles and applications*. New York: Irvington.

Flores, M. M. (2010). Using the concrete-representational-abstract sequence to teach subtraction with regrouping to students at risk for failure. *Remedial and Special Education, 31*(3), 195–207. doi:10.1177/0741932508327467.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Math flash* [computer software]. Nashville: Vanderbilt University Press.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493–513. doi:10.1037/0022-0663.97.3.493.

Fuchs, L. S., Fuchs, D., & Hollenbeck, K. N. (2007). Extending responsiveness to intervention to mathematics at first and third grades. *Learning Disabilities Research and Practice, 22*(1), 13–24. doi:10.1111/j.1540-5826.2007.00227.x.

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., et al. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disabilities Quarterly, 31,* 79–92.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552. doi:10.1037/a0025510.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345–362. doi:10.1037/0033-2909.114.2.345.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*(2), 121–151. doi:10.1016/j.jecp.2004.03.002.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The number sets test. *Journal of Psychoeducational Assessment, 27*(3), 265–279. doi:10.1177/0734282908330592.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33,* 18–28.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38,* 93–304.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242. doi:10.3102/0034654309334431.

Gickling, E. E., & Havertape, S. (1981). *Curriculum-based assessment (CBA)*. Minneapolis: School Psychology Inservice Training Network.

Gickling, E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 205–218.

Goldman, S. R. (1989). Strategy instruction in mathematics. *Learning Disability Quarterly, 12,* 43–55.

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional technology: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23–40). Columbus: Merrill.

Hasselbring, T. S., Goin, L. S., & Bransford, J. D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. *Focus on Exceptional Children, 20*(6), 1–7.

Hasselbring, T. S., Lott, A., & Zydney, J. (2006). *Technology-supported math instruction for students with disabilities*: *Two decades of research and development*. http://www.ldonline.org/article/6291/. Accessed 4 Dec 2012.

Jitendra, A. K. (2007). *Solving math word problems*. *Teaching students with learning disabilities using schema-based instruction*. Austin: PRO-ED.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47,* 1475–1490.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*(2), 103–119. doi:10.1016/S0022-0965(03)00032-8.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850–867. doi:10.1037/a0014939.

Kroeger, S. D., & Kouche, B. (2006). Using peer-assisted learning strategies to increase response to intervention in inclusive middle math settings. (Cover story). *Teaching Exceptional Children, 38*(5), 6–13.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics intervention for children with special needs: A meta-analysis. *Remedial and Special Education, 24,* 97–114.

Lin, F., & Kubina, R. M., Jr. (2005). A preliminary investigation of the relationship between fluency and application for multiplication. *Journal of Behavioral Education, 14,* 73–78.

Martens, B. K., & Eckert, T. L. (2007). The instructional hierarchy as a model of stimulus control over student and teacher behavior: We're close but are we close enough? *Journal of Behavioral Education, 16,* 83–91.

Mazzocco, M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research and Practice, 20*(3), 142–155.

Meichenbaum, D. H., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. *Journal of Abnormal Psychology, 77,* 115–126.

Mercer, C. D., & Miller, S. P. (1992). Teaching students with learning problems in math to acquire, understand, and apply basic math facts. *RASE*: *Remedial and Special Education, 13*(3), 19–35. doi:10.1177/074193259201300303.

Methe, S. A., Hojnoski, R., Clarke, B., Owens, B. B., Lilley, P. K., Politylo, B. C., & Marcotte, A. M. (2011). Innovations and future directions for early numeracy curriculum-based measurement: Commentary on the special series. *Assessment for Effective Intervention, 34*(6), 200–209. doi:10.1177/1534508411414154.

Miller, S., & Mercer, C. D. (1993). Mnemonics: Enhancing the math performance of students with learning difficulties. *Intervention in School and Clinic, 29*(2), 78–82.

Montague, M. (2008). Self-regulation strategies to improve mathematical problem solving for students with learning disabilities. *Learning Disability Quarterly, 31*(1), 37–44.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Kindergarten predictors of recurring externalizing and internalizing psychopathology in the third and fifth grades. *Journal of Emotional and Behavioral Disorders, 17*(2), 67–79.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. Reston: National Council of Teachers of Mathematics.

National Governers Associaton Center for Best Practices, Council of Cheif State School Officers. (2010). *Common Core state standards for mathematics*. http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf. Accessed 3 Dec 2012.

National Mathematics Advisory Panel. (2008, March). Foundations for success: *The Final Report of the National Mathematics Advisory Panel*, U.S. Department of Education: Washington, DC. http://www.colonialsd.org/colonial/lib/colonial/national_math_panel_highlights_march_2008.doc. Accessed 11 Nov 2008.

Poncy, B. C., Skinner, C. H., & O'Mara, T. (2006). Detect, practice, and repair: The effects of a classwide intervention on elementary students' math-fact fluency. *Journal of Evidence-Based Practices for Schools, 7*(1), 47–68.

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learning Disabilities Research and Practice, 24*(1), 1–11. doi:10.1111/j.1540-5826.2008.01272.x.

Rivera, D. M., & Bryant, B. R. (1992). Mathematics instruction for students with special needs. *Intervention in School and Clinic, 28,* 71–86.

Shapiro, E. (2011). *Academic kills problems: Direct assessment and intervention*. New York: The Guilford Press.

Skinner, C. H., & Daly, E. (2010). Improving generalization of academic skills: Commentary on the special series. *Journal of Behavioral Education, 19*(1), 106–115.

Skinner, C. H., Belfiore, P. J., Mace, H. W., Williams-Wilson, S., & Johns, G. A. (1997). Altering response topography to increase response efficiency and learning rates. *School Psychology Quarterly, 12*(1), 54–64. doi:10.1037/h0088947.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428–444. doi:10.1111/j.1467-8624.2004.00684.x.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games-but not circular ones-improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545–560. doi:10.1037/a0014239.

Swanson, H. (2009). Science-supported math instruction for children with math difficulties: Converting a meta-analysis to practice. In S. Rosenfield & V. Berninger (Eds.), *Implementing evidence-based academic interventions in school settings* (pp. 85–106). New York: Oxford University Press.

Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject design intervention research for students with learning disabilities. *Journal of Learning Disabilities, 33,* 114–136.

Thurber, R., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31*(4), 498–513.

Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities, 36*(5), 449–458.

VanDerHeyden, A., & Burns, M. (2008). Using curriculum-based assessment and curium-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30,* 15–31.

VanDerHeyden, A. M., & Witt, J. C. (2008). Best practices in can't do/won't do assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V: Vol. 2 data-based decision making* (pp. 195–208). Washington, DC: National Association of School Psychologists.

Woodward, J. (2006). Developing automaticity in multiplication facts: Integrating strategy instruction with timed practice drills. *Learning Disability Quarterly, 29,* 269–289.

Tier 3 Primary Grade Reading Interventions: Can We Distinguish Necessary from Sufficient?

Stephanie Al Otaiba, Jill Allor, Miriam Ortiz, Luana Greulich, Jeanie Wanzek and Joseph Torgesen

Since publication of the previous version of this book, the implementation of response to intervention (RTI) has become more widespread. All 50 states now encourage RTI for prevention purposes, and a growing number of states allow it for the identification of learning disabilities (Fuchs and Vaughn 2012; Zirkel and Thomas 2010). A review of the evidence base for RTI (Gersten et al. 2008) identified five recommendations for practice, which have become the five core components of RTI that are rapidly gaining traction in state laws or guidelines (Berkeley et al. 2009; Zirkel and Thomas 2010). These core components include universal screening, a high-quality core reading program, progress monitoring, increasingly intensive tiers of intervention, and fidelity of implementation.

The purpose of this chapter is to describe the most intensive tier of intervention within RTI (tier 3) that can be implemented for reading difficulties with primary grade students (in kindergarten through third grade). Within multi-tier RTI systems, tier 3 is reserved for the very few students with the most persistent reading difficulties, having demonstrated significant difficulties in learning to read even with tier 1 and tier 2 supports. As explained in greater detail, tier 3 may be a last layer of intervention in general education that serves as part of a prereferral process or it may be special education. In this chapter, the authors consider the intensity of tier 3 interventions to be appropriate for students who are struggling, students with reading or mild-to-moderate intellectual disabilities.

In this chapter, first, a conceptual framework is provided. Second, sources of variability in tier 3 implementation within the USA are described. Third, relevant research and evidence from a literature review of multi-tier investigations that have provided tier 3 for students who did not show adequate reading progress in prior tier 1 instruction and tier 2 intervention are summarized. A summary table provides an overview of the research, delineating intensity of intervention including the tiers, the interventionists, the grade levels in which tiers were delivered, the number of hours and sessions per week, the grouping ratio, the diversity of participants, and the intervention components (see Table 1). Advances in brief experimental analyses within problem-solving, or problem analysis, models of RTI are described. Fourth, because a review of all special education interventions is beyond the scope of this chapter, a successful multi-year study is described to give readers an example of

S. Al Otaiba (✉) · J. Allor · M. Ortiz
Southern Methodist University, Dallas, USA
e-mail: salotaiba@smu.edu

L. Greulich
Andrews University, Berrien Springs, USA

J. Wanzek · J. Torgesen
Florida Center for Reading Research, Florida State University, Tallahassee, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_23

Table 1 Studies examining intensity and effects of tier 3

Study	Tier	Interventionist and training	Grades	Minutes of instruction; sessions per week; duration	Grouping ratio	Diversity of participants and school setting	Intervention components
O'Connor et al. 2005	Tier 1	Research assistants; well trained	Kindergarten—3rd grade	Not reported	–	Diverse SES; diverse participants;	Professional development focusing on findings of National Reading Panel 2000
	Tier 2		Kindergarten—3rd grade	K: 10–15 min 3 days per week 1st—3rd: 20–25 min; 3 days per week	2–3 students		Segmentation of one syllable spoken words; Letter knowledge, sound knowledge; *Curriculums*: Bob Books, Reading Mastery I, Open Court
	Tier 3		1st–3rd grade	30 min daily	1–2 students		Same as tier 2
Vaughn et al. 2009	Tier 1	Research assistants; well trained		Not reported	–	Diverse participants	Professional development
	Tier 2		1st grade	30 min daily; 13 or 26 weeks	4–6 students		Phonics; word recognition; fluency; passage reading; passage comprehension
	Tier 3		2nd grade	50 min daily; 26 weeks	2–4 students		Letter sound review; phonics; word recognition; vocabulary; fluency; passage reading: passage comprehension
Wanzek and Vaughn 2008, 2011	Tier 1	Graduate and research assistants; well trained	1st grade	Nr		Low SES; Title I, mainly minority	Professional development; approximately 14 sessions
	Tier 2 (2008)		1st grade	30 min daily, 13 weeks 60 min daily, 13 weeks	4–5 students		Phonics; word recognition; fluency; passage reading; passage comprehension

Table 1 (continued)

Study	Tier	Interventionist and training	Grades	Minutes of instruction; sessions per week; duration	Grouping ratio	Diversity of participants and school setting	Intervention components
	Tier 3 (2011)		2nd and 3rd grade	2nd grade: 50 min daily; 26 weeks 3rd grade: 50 min daily; 26 weeks	2nd grade: Fall 2–4 students Spring: 1 student 3rd grade: 1 student		Phonemic awareness; phonics; word recognition; fluency; vocabulary; text reading; comprehension of decodable text
(Tier 1 and 2) Mathes et al. 2005 (Tier 3) Denton et al. 2006	Tier 1	Certified teachers; well trained	1st grade	Nr		Diverse SES; Diverse participants	Professional development; access to assessment and progress monitoring data. Information on providing differentiated instruction and peer tutoring
	Tier 2		1st grade	40 min daily October to May	3 students		*Proactive reading:* Word ID; letter-sound correspondence; sounding out and reading words; spelling words in isolation; immediate positive feedback *Responsive reading:* phonemic awareness; phonemic decoding; fluency building; assessment; letter and word work; supported reading; supported writing
	Tier 3		1st grade (Remedial), 2nd, and 3rd grade	Two 50-min sessions with a 10-min break; 8 weeks 1 h daily 8 weeks	2 students		1st 8 Weeks: Phono-Graphix 2nd 8 Weeks: Read Naturally

Table 1 (continued)

Study	Tier	Interventionist and training	Grades	Minutes of instruction; sessions per week; duration	Grouping ratio	Diversity of participants and school setting	Intervention components
Vellutino et al. 2006	Tier1	Certified teacher; well trained	Not reported	Not reported		Middle SES; Rural New York	
	Tier 2		kindergarten	30 min; two times per week	2–3 students		Print concepts; letter ID; phonological awareness; letter-sound mapping; sight word learning; guided reading; story reading
	Tier 3		1st grade	Nr	1 student		Phonological skills; text processing; sight word ID; writing
Vaughn et al. 2003	Tier1	Hired tutors with bachelor's degrees	2nd grade	Nr		Mostly hispanic/Latino; 33 % English Language Learners	
	Tier 2			35 min daily; 10 weeks 35 min daily; 10 weeks	3 students		Phonemic awareness; phonics; fluency; instructional level reading; comprehension; spelling
	Tier 3			35 min daily; 10 weeks	Nr		Increased word study and fluency; reduced phonological awareness

SES socioeconomic status, *ID* intellectual disabilities
Phono-Graphix, (McGuiness et al. 1996); Read Naturally, (Ihnot et al. 2001); Open Court Reading, (Adams et al. 2000); Reading Mastery Fast, (Engelmann and Bruner 1995); National Reading Panel 2000; Bob Books, (Maslen and Maslen 1987); Open Court, (SRA/McGraw-Hill 1998)

how intensive and extensive tier 3 reading intervention must be to help students with low IQs (including those with mild-to-moderate intellectual disabilities). Finally, implications for practice and identify directions for future research are discussed.

Conceptual Framework

As in our chapter, in the previous edition of this book (Al Otaiba and Torgesen 2007), the first part of our conceptual framework is knowledge about reading development (e.g., Snow et al. 1998) and the content of the reading intervention that supports learning (e.g., National Early Literacy Panel 2009; National Reading Panel 2000). From this work, and from the Simple View of Reading (Gough and Tunmer 1986), it is known that two stumbling blocks impede the pathway to becoming a proficient reader: Problems in understanding and using the alphabetic principle to acquire fluent and automatic word reading skills; and problems in acquiring the verbal knowledge and strategies that are specifically needed for comprehension of written material. The purpose of RTI is to prevent children from lagging behind in the development of fluent word reading and vocabulary/strategic thinking skills because each of these areas is critical for proficient reading comprehension (Snow 2002). Furthermore, once problems develop in either of these areas, motivation and enjoyment of reading can suffer. Therefore, in order to remediate serious reading difficulties, tier 3 interventions must be intensive enough to powerfully accelerate development in word reading skills (phonemic awareness, phonics, fluency), while also supporting the development of vocabulary, conceptual knowledge, reading comprehension, and thinking skills.

Thus, the second aspect of the conceptual framework of this chapter involves the most important dimensions of instructional method that are characteristic of successful tier 3 interventions designed for the small proportion of children who will need very intensive and formative, data-guided, individualized intervention. The successful interventions described in this chapter invariably increase the intensity and the explicitness of intervention in Tier 3 and also increase the frequency of formative progress monitoring to test the efficacy of the intervention and to determine when changes are needed.

Mellard et al. (2009) expanded the construct of intensity to include 10 variables. The first three are considered dosage related and include minutes of intervention, frequency, and duration. Other variables are group size, immediacy of feedback, support for motivation, level of mastery required, opportunities to respond, frequency of transitions, specificity of goals, and interventionist skill and training. Wanzek and Vaughn (2007) examined the research on extensive (defined as over 100 sessions) early literacy interventions and reported high reading achievement for students who were provided such intensive intervention. Effects were stronger for younger (K–first grade) than older (second to third grade) students. Although the authors could not categorize the 18 studies by tier of intervention, intervention for 10 of the studies provided one-on-one support for 5 months–2.5 years. Although their findings are positive, only 2 of the 18 studies described the effect of intervention provided to students who had not responded to previous intervention and only one delivered tier 3 intervention (Vaughn et al. 2003). A subsequent synthesis of extensive interventions beyond grade 3 noted much smaller effects for students with reading difficulties further highlighting the importance of early literacy interventions, including intensive interventions (Wanzek et al. 2012).

Of course, tier 3 interventions must also address motivational or behavior management issues, given that it is frequently an area of challenge for students who struggle in reading. Students struggling to read experience decreased motivation for reading (Morgan et al. 2008; Wigfield and Guthrie 1997), and lose opportunities for development of increasingly sophisticated reading strategies (Brown et al. 1986). Thus, even the most powerful remedial interventions are not likely able to help students who fall behind and close the gap in reading with students who are learning to read normally (Torgesen 2005).

Sources of Variability in Tier 3 Practice

Although RTI has been used for over two decades, there remains considerable variability about how tier 3 is operationalized across, and even within states (Jenkins et al. 2013; Mellard et al. 2010; Mellard et al. 2009). One source of this variability is related to the model or approach used; some states are implementing a standard protocol approach for intervention, others use problem-solving approaches, and others still use a hybrid, or combination approach (Berkeley et al. 2009). Briefly, in a standard approach, all students with similar needs receive the same evidence-based intervention. Sometimes, interventions are implemented by teachers who received professional development, but more frequently these interventions have been delivered by highly trained research staff. Fidelity of the treatment is carefully observed. In contrast, problem-solving approaches have tended to be more individualized and generally include the following steps: problem is identified, gap between current and expected performance is documented, problem solutions are identified (e.g., an intervention such as modeling or praise or goal setting), the intervention is delivered, and progress is monitored. If progress is insufficient, then another cycle of problem identification is begun; if progress is sufficient, the process is ended. A hybrid approach might begin with standard protocols, but if these are unsuccessful, a teacher or an RTI team would move toward a more individualized problem-solving approach.

For a more thorough discussion of these models, see D. Fuchs et al. 2003 and for a meta-analytic review of these models, see Burns et al. 2005. In their review, Burns et al. focused on field-based versus researcher-implemented RTI, but noted that the field-based approaches followed the problem-solving approaches and that the researchers followed standard protocol approaches. Encouragingly, the effect size for student outcomes was 1.14 for the researcher-implemented standard protocol models and 0.94 for the field-based problem-solving models. No studies in their meta-analysis directly compared these two approaches to determine their relative impact on student outcomes or on student referral and placement in special education.

Another variable aspect of tier 3 is that in some states it is within general education and in other states, it is part of special education (Zirkel and Thomas 2010). Researchers themselves are divided as to whether tier 3 should be special education or part of general education or, in fact, how many tiers there should be (Fuchs et al. 2003). For example, Fuchs et al. argued that tier 3 should include special education to ensure that the most intensive level of intervention is provided to the students with the most persistent needs and that such intervention is provided by highly skilled educators who can use formative data to provide individualized intervention (cf. Deno and Mirkin 1977; Fuchs et al. 2010; Fuchs and Fuchs 1986). Another interpretation is the possibility that tier 3 is a part of general education implemented prior to evaluation for possible special education services (Gersten et al. 2008). For the purposes of this chapter, the authors consider tier 3 to include tier 3 and anything even more intensive, but a thorough review of special education remedial interventions is beyond the scope.

It is also important to clarify that tier 3 is rarely delivered as a stand-alone intervention, but is more typically delivered in addition to tier 1 or tier 1 and tier 2. In fact, in a survey of 41 model school sites implementing RTI, only a few schools reported that the dosage of tier 3 was in fact more intensive than tier 2 (Mellard et al. 2010). A more recent survey of 62 schools who were implementing RTI found that 87% of respondents indicated that tier 3 provided more minutes of instruction than tier 2, but expressed concern about the challenge of interpreting time as some schools layered tier 3 on top of tier 2 (and/or tier 1; Jenkins et al. in press). The same could be said for special education, in that Jenkins et al. reported that a majority of students receiving special education in reading participated in tier 1 with accommodations and modifications and with supplemental intervention in a variety of models.

A final source of variability is the relative lack of research evidence for tier 3 relative to the strong evidence base for tier 2 (Gersten et al. 2008), which is potentially problematic, given that RTI implementation is increasing dramatically. Specifically, Gersten et al. (2008) expressed concern for the low level of evidence for tier 3 for those students who did not respond to evidence-based tier 1 and tier 2. Nevertheless, the guide recommends that tier 3 "provide intensive intervention on a daily basis that promotes the development of the various components of reading proficiency to students who show minimal progress after a reasonable time in tier 2 small group instruction" (p. 6). Moreover, Gersten et al. suggested that tier 3 interventions should be delivered by qualified and well-trained teachers. Further suggestions include narrowing the focus of intervention, adjusting pacing to ensure mastery, minimizing the group size to provide more time on task and extensive practice with feedback, individualizing informed by frequently collected formative data, providing intervention daily, and extending the dosage and duration of intervention as needed based on data.

Relevant Research and Evidence of Effectiveness of Interventions

This section describes the nature and results of multi-tier interventions to provide an understanding of the intensity of instructional conditions that need to be in place to prevent and remediate early reading problems. We summarize evidence from a literature review of the existing empirical multi-tier investigations that have provided tier 3 for students who did not show adequate reading progress in prior tier 1 instruction and tier 2 intervention.

There are a limited number of studies (or combination of studies) in peer-reviewed journals that yield information about the effects of tier 3 provided to students who did not respond to tiers 1 and 2 (Denton et al. 2006; O'Connor et al. 2005; Vaughn et al. 2003, 2008, 2007, 2009; Wanzek and Vaughn 2010; Vellutino et al. 2006). The aforementioned studies were all researcher delivered and used standard-protocol approaches. Table 1 provides an overview of the research, delineating intensity of intervention including the tiers, the interventionists, the grade levels in which tiers were delivered, the number of hours and sessions per week, the grouping ratio, the diversity of participants, and the intervention components. For the ease for readers, the table follows the order in which the studies are introduced in text. The authors also more briefly describe implications from other studies they located, including a single-case design involving a problem-solving RTI approach (VanDerHeyden et al. 2007), a quasi-experimental study examining professional development to support school-implemented RTI that included a historical baseline year (Harn et al. 2011), and a case study (Orosco and Klingner 2010).

Perhaps the earliest multi-tier study was conducted by O'Connor et al. (2005) and followed children from kindergarten through third grade. Tier 1 involved professional development for teachers to improve their core reading program. Tier 2, involving small group phonics intervention, was implemented beginning in January of kindergarten and continuing throughout first grade. By January of first grade, tier 3 began for students who did not meet criteria for responsiveness. Tier 2 and 3 then continued until grade 3, with children moving in and out of tiers as needed. At the end of the study, 40 % of students in tier 3 were at grade level for both word reading and fluency. Because movement and amounts of intervention are complex, findings from this study are complicated to interpret. However, one straightforward implication is the complexity of managing RTI as teachers must know how to use data to move students and schools must have adequate resources.

Two research teams have carefully implemented a series of studies and their findings document that for students who do not benefit from tier 1 and 2, tier 3 will likely need to be intensive, data driven, and extensive (defined by Wanzek and Vaughn 2007 as lasting at least 100 sessions). Both research teams provided 100 additional hours of tier 3 to students who did not benefit from a combination of supported tier 1 (professional development, universal screening, and an evidence-based first-grade core reading program) and researcher-delivered tier 2. The first team, Vaughn et al. (2008, 2007, 2009; Wanzek and Vaughn 2008), validated that students whose teachers received professional development and support for tier 1 achieved higher reading performance than students in control classrooms.

Vaughn et al. found that supported tier 1 led to fewer students at risk. Further, tier 2 (conducted for between 13 and 26 weeks across first grade) helped a majority of first graders read on grade level. Tier 3 intervention, provided to first graders with insufficient response to tier I and tier 2 interventions, that focused on increasing instructional time in intervention (two 30-min sessions per day) with little change to the type of intervention provided in tier 2 was not sufficient for significantly improving student reading outcomes. However, when tier 3 intervention that included a change in instruction as well as smaller group size and increased instructional time, was provided to a small number of second-grade students who had not benefited from tier 1 and 2, these students read fewer than 27 words correct per minute at the start of second grade. Intervention included 100 daily 50-min small group researcher-provided sessions of explicit and systematic phonics, word recognition, fluency, and comprehension. Students in tier 3 showed a steeper slope in reading progress than students who did not receive tier 3. Even with this intensity, when scores of oral reading fluency, nonsense word decoding, word reading, and comprehension were examined, significant differences favored students who received tier 3 only on untimed word reading and comprehension. An important implication of this series of studies is that a number of students in tier 3 have persistent and extensive needs and require (a) different interventions or (b) ongoing intensive interventions that are guided by formative assessment.

The second team conducted two studies, the first by Mathes et al. 2005 and a follow up by Denton et al. (2006), who worked with two groups of students. The first group was similar to nonresponders in the just-described series of studies conducted by Vaughn et al. These were students in second grade who had not responded to a successful year-long tier 2 intervention including teacher-supported tier 1 (Mathes et al. 2005). Notably, in the Mathes et al study, all but 3 % of students who scored below the 30th percentile on the WJ-III could read on grade level. To increase the number of students, a second group of students were added who included schoolmates of the nonresponders to Mathes et al.; they were in second and third grade with similarly low scores. Thus, the students had not responded to prior school-delivered tier 1 and 2. Denton et al. (2006) found large individual differences in response, according to various standardized measures of reading, with some children demonstrating little or no growth. An important implication of this study was that tier 3 (or special education interventions) for students with the most significant and intractable reading problems will likely require something more than a standard-protocol approach, and may require extended time, perhaps over the course of several years.

There is evidence that even when students succeed in tier 3, students who were the most difficult to remediate will likely need ongoing intensive support. Vellutino et al. (2006) followed students to third grade who had participated in an empirical study of a three-tier model conducted over kindergarten and first grade. Students at risk for reading difficulties were identified in kindergarten and were assigned randomly to receive tier 2 small group (2–3 children) intervention for 30 min twice a week provided by a research team or business-as-usual support from schools. Effect sizes favored the treated students at the end of kindergarten relative to the business-as-usual comparison; for phonological skills, these ranged from 0.09 for rhyme to 1.66 for phonemic segmentation. For phonics skills, the effect sizes ranged from 0.51 on a standardized published measure to 1.30 on a researcher-made decoding task.

At the start of first grade, students still exhibiting difficulties in understanding or applying the alphabetic principle were randomly assigned by Vellutino et al. (2006) to daily one-to-one tutoring or to business-as-usual support from schools. Thereafter, the focus of the analysis was to examine growth of subgroups of children who participated in treatment who were (a) no longer at risk ($n = 81$), (b) less difficult to remediate in first grade ($n=26$), and (c) the most difficult to remediate ($n=19$). Students who were most difficult to remediate failed to maintain the gains made after receiving both tier 2 (in kindergarten) and tier 3 (in first grade), and their word identification standard scores decreased from 94.63 immediately following treatment at the end of first grade to 79.16 at the end of third grade. A similar pattern of declining performance occurred on word attack and comprehension measures. Students who were less difficult to remediate across tier 3 decreased by only 1–6 standard scores on word attack and word identification, but increased 2 points on comprehension. By contrast, students whose reading problems were successfully prevented by kindergarten tier 2 showed a trend of increasing standard score points on all three measures. An important implication of this study is the need to learn how schools can ensure that students receiving tier 3 in the early grades continue to receive support as long as they need it.

At least one study provides promising evidence for English language learners (ELLs) in multi-tiered systems but, it appears that more research is needed to adapt models that are English only to be more culturally and linguistically responsive. Vaughn et al. 2003 recruited 45 second graders at risk for reading problems; a majority were in Title 1 schools and were ELLs. Students who did not make exit criteria after 20 weeks of tier 2 were provided with even more intensive and individualized intervention (tier 3). Students who scored below 5 on the Texas Primary Reading Inventory (Texas Education Agency 1998) were identified for tier 3, and their skills were assessed weekly using 1-min informal dictation exercises to determine mastery of sight words and phonics patterns. Intervention was more intense and focused on decoding and fluency. A total of 24 students made exit criteria. Of these, 16 were able to improve by at least one word per week in tier 1 and the remaining 8 were unable to make even this minimal progress. By contrast, 11 students never made exit criteria. Whether or not students received bilingual education or ESL instruction did not appear related to success. An encouraging implication of their research is that some students will benefit from relatively small amounts of intervention, but another implication is consistent with concerns from the study by Vellutino et al. (2006) suggesting that benefits may fade without additional support.

Other concerns about RTI models for ELLs are consistent with prior studies by Klingner et al. (Hoover et al. 2012; Klingner and Harry 2006), who have described the difficulties many school personnel encounter in distinguishing who needs intervention from who has a lack of English proficiency. They also noted the challenge of linking data from multiple resources (and across languages) to make accurate instructional decisions. In a case study observing RTI (field based, not researcher delivered) in a district with a high number of ELLs, Orosco and Klingner (2010) found that the success of RTI for ELLs was particularly dependent on teacher knowledge, professional development, curriculum, and resources. They reported that if a model of RTI follows a standard protocol that has evidence only for English speakers and does not take into consideration the needs of ELLs, students' response was diminished. Thus, they recommend that "a one-size-fits-all policy approach to RTI might not work" (p. 283). Further, their findings support the need for future research to include mixed methods and qualitative studies to explicate the contextual variables to support RTI.

There is also evidence from two additional studies that without support, teachers are not likely to use data to differentiate in tier 1 and tier 2, and that tier 3 may involve multiple efforts that lack alignment. An example of the need to help schools align tier 3 with tiers 1 and 2 and to coordinate resources is Harn et al. (2011). They conducted a quasi-experimental study to support field implementation of RTI in a high-need district. In year 1, or the baseline year, they observed that neither tier 1 nor tier 2 was differentiated, and there was no tier 3 support. Further, struggling students received as many as five different programs for varying times by different personnel. It was challenging for schools to collaborate resources across Title 1, special education, and services for ELLs. In the second year, Harn et al. provided professional development to "not only show people how the interventions were to be delivered but also to build knowledge and understanding on why high levels of student engagement, provision of corrective feedback and use of consistent instructional language were important to student learning" (p. 351). The year 2 support for coordinating instruction led to higher student outcomes and increased opportunities to respond within tier 2 relative to baseline.

To better understand how teachers use data to inform instruction within a multi-tier model, VanDerHeyden et al. conducted a multiple baseline study over 4 years in five schools. They used the problem-solving model, the System to Enhance Educational Performance that was developed by Witt et al. (STEEP; Witt et al. 2000). The first component of STEEP was school-wide universal screening (with 15 % of students screened served in tier 2). In tier 2, teachers provided extra instruction and rewarded students for increasing previous performance; students who did not improve (11 %) were moved to tier 3, also provided by classroom general education teachers. STEEP data usage was related to a decline in the number of special education referrals and improved accuracy of identification. However, teachers found it challenging to use formative data to make decisions independently, so the school psychologist was instrumental, at least at this stage of implementation. An important implication is the need for both teacher preservice and professional development to support teacher formative data use.

Some converging evidence suggests that it may be possible (and effective) to identify students who will require tier 3 without requiring that they progress through tier 2. Al Otaiba et al. (2014) compared two models of RTI for first grade reading; in Typical RTI, students with the weakest skills progress to Tier 3 after not responding to Tiers 1 and 2. In the Dynamic RTI condition, such students could begin immediately in tier 3. Controlling for fall scores, findings revealed statistically significant reading outcomes (ES = .36) favoring students in the Dynamic condition. However, heterogeneity between stronger and weaker responders within tier 3 became more pronounced as the interventions progressed, and students who were ready to exit tier 3 intervention (because they scored above the 40th percentile on screeners and had made adequate growth on the screener) still could not read at the level of text complexity as their tier 1 classmates (Al Otaiba et al. 2014). Thus, it was challenging to transition students from higher to lower levels of intervention, which suggests the need to better match student needs to resources.

Once the promise of standard reading intervention protocols are exhausted, interventionists may have to turn to more in-depth assessment and analysis to determine the appropriate intervention. Brief experimental analyses (BEA) of academic skills may embody what D. Fuchs et al. (2012) referred to as "Smart RTI" (p. 263). There is a rich history of using BEA to help select behavioral and academic interventions (e.g., Daly et al. 1997; Johnston and Pennypacker 1993; Sidman 1960). BEAs do not assume that one intervention will work for all children, but rather allow teachers to test (often in an hour or less) what intervention will likely work (or not) for individual children (see McComas chapter for more information). For example, a teacher may add reinforcement to see if praise or a combination of praise and self-graphing increases sight word fluency to rule out a performance, rather than a skill deficit (or more commonly in the teaching lexicon to learn whether the problem is a *won't do* or a *can't do* issue). Another possibility is that a teacher might be concerned that a student is reading too few words per minute because the reading material is beyond her independent reading level (cf. Lentz 1988). If so, the teacher could briefly test if reading easier material or reading a text with more known words leads to faster oral reading fluency.

To date, most research on BEAs has focused on improving oral reading fluency (e.g., Daly and Martens 1994; Daly et al.1997, 1999; Noell et al. 2001), but recently it has also been used with earlier reading skills with kindergarteners (Petursdottir et al. 2009). Burns and Wagner (2008) conducted a meta-analytic review of 13 studies that focused on oral reading fluency and reported that BEAs were effective at identifying which interventions would lead to the largest effect sizes. Ensuring that children read at their instructional level also led to higher effect sizes. Burns and Wagner reported that interventions identified through BEA led to an average fluency increase of 73 % over the median baseline score (or roughly 30 words correct per minute). A variety of interventions were effective at improving oral reading fluency, with the largest effects related to the combination of listening to passages, repeated reading, and performance feedback with or without incentives. Burns and Wagner concluded that BEA was a promising technique for linking data to intervention. However, to date, BEAs have mostly been conducted by school psychologists rather than teachers. Future research is needed to train general and special educators to use these approaches.

Tier 3 Instruction for Students with Intellectual Disabilities

Because a thorough review of special education intervention is beyond the scope of this chapter, the readers are referred to meta-analyses and literature reviews of remedial interventions provided to students with learning disabilities (e.g., Swanson et al. 1999) and to students with intellectual disabilities (e.g., Browder et al. 2006). This section of the chapter provides a description of an example of a very intensive multi-year tier 3 reading intervention that was delivered to students with low IQs (including students with mild-to-moderate intellectual disabilities—ID).

Recent research supports that students with ID respond favorably to comprehensive, explicit, and systematic reading instruction when it is delivered with intensity and instructional decisions are individualized and based on progress monitoring data (e.g., Allor et al. 2010b; Lemons and Fuchs 2010). Although generally students with lower IQs respond more slowly to instruction, they are responsive to the same type of reading instruction (i.e., scientifically based reading instruction) that has a solid history of effectiveness with students with higher IQs, including those with a learning disability (LD). With explicit and intense instruction, they are able to develop reading skills similar to more typically developing peers, in that they process the internal structure of words and are able to comprehend at levels commensurate with their general language comprehension (Allor et al. 2014, 2010c). In a recent longitudinal study, elementary students with IQs ranging from 40 to 80 were followed for up to 4 years with very positive findings (Allor et al. 2012; for earlier reports on a subset of these participants, see Allor et al. 2010a, b). Participants were randomly assigned (within IQ range and school) to either a contrast "business-as-usual" group or a treatment group that received consistent, intensive scientifically based reading instruction for up to 4 years. Statistically, significant differences were found across all IQ levels on a variety of reading measures, including measures of phonological awareness, phonological decoding, word recognition, fluency, and comprehension. Specifically, students in the mild ID range (i.e., 55–69) demonstrated clear growth on progress-monitoring measures.

Students from the Allor et al. (2014) study in the mild ID range demonstrated mastery on a phoneme segmentation fluency (i.e., 35 segments per minute) after approximately 75 weeks of instruction (i.e., approximately 2 academic years). Similarly, on nonsense word fluency, students demonstrated mastery (i.e., 50 letter sounds per minute) after approximately 50 weeks of instruction (i.e., approximately 1½ academic years). On first-grade oral reading fluency passages, students began the study reading approximately ten words per minute (wpm) and reached the ending first-grade-level benchmark of 40 wpm after 60 weeks of instruction (i.e., nearly 2 academic years) and were reading 70 wpm (also on first-grade passages) by the end of the study (i.e., approximately 4 academic years). In sum, progress was slow, but, by the end of the study, students with mild ID were reading ending first-grade-level passages with fluency and basic comprehension.

These findings are encouraging as they demonstrate that students with mild intellectual disabilities respond to explicit and systematic instruction, yet they also demonstrate that this response is relatively slow and requires intense instruction in terms of fidelity of implementation and consistent instruction over a long period of time. The students in the Allor et al. (2014) study all received at least 40–50 min of daily small group (and sometimes individual) instruction by a highly trained research intervention teacher. For some students, this was their only reading instruction, but others participated in additional instruction with their current special education and/or general education teacher (typical instruction provided by the teacher). Also, in the last year of this study, additional practice was provided by a variety of people, including family members, volunteers, or peer tutors.

Ideally, students would receive their primary reading instruction in a small group setting with a highly trained teacher and also participate in additional reading instruction. For students with more moderate-to-severe IDs, we anticipate that this primary instruction would most often occur in a special education setting and students would also participate in tier 1 instruction with modifications so that instruction would include appropriate practice at their instructional level. It is also possible that other small group instruction by literacy specialists might also be appropriate for these students. The key is that these instructional placements should be based on student need and ongoing progress monitoring. The type of instruction employed in this and other studies is consistent with the same type of instruction required for students with LD; however, individualization was required (Allor et al. 2010c; Allor and Mathes 2012; Mathes and Torgesen 2005). Obviously, repetition of lessons and intensive practice was key, but other modifications were required as well, particularly behavior modification techniques and directly teaching students to transfer skills from one task to another. For example, students may have successfully sounded out words in familiar lessons, but needed explicit modeling to apply those skills to words in context. Teaching students with ID to read will require the expertise of special education teachers familiar with the unique needs of this population, as well as the expertise of literacy specialists.

Implications for Practice and Identifying Areas for Future Research

As schools scale up RTI, it is timely and critical to learn whether the five core RTI components recommended by Gersten et al (2008) (i.e., evidence based core program, screening, progress monitoring, increasingly intensive intervention and fidelity of implementation) are indeed a necessary and sufficient set of best practices. Converging evidence indicates that persistently weak responders exist, and they will likely require a hybrid approach as standard protocols are exhausted and students continue to require ongoing intensive interventions that are guided by formative assessment. The review of the research indicates studies have met Mellard et al.'s (2010) criteria for intensity and Wanzek and Vaughn's (2007) criteria for being extensive. Even if more sophisticated two-stage screening and BEAs are promising for matching resources to student needs, it seems likely that teachers will need help to use data confidently to change interventions (Harn et al. 2011; VanderHeyden et al. 2007). Further, we agree with Orosco and Klingner (2010) that it is unlikely that any single model of RTI will work for all schools.

Nevertheless, it is concerning that school systems, and special educators, lack guidance from research about how to implement tier 3 and that they have such variable guidance about tier 3 from existing RTI law and policy (Berkeley et al. 2009; Zirkel and Thomas 2010a, b). Concerns about this variability are echoed in recent surveys of special education directors (Werts et al. 2009) and in surveys about RTI implementation (Jenkins et al. in press; Mellard et al. 2009). Moreover, most surveys and observations about RTI implementation have occurred only in schools that have been nominated as "high performing" or who have been implementing RTI for some time. Now that some core components are being widely used (e.g., uniform screening), how the data are used may be more important than whether a school collects the data in explaining student outcomes.

This lack of real-world information about tier 3 RTI implementation is problematic, given that RTI is being implemented in all 50 states as a means of preventing reading disabilities and that it is increasingly being used as a means of identifying reading disabilities that are not due to poor instruction. It is vital to learn the extent to which specific tier 3 implementation features are associated with stronger student outcomes for students receiving tier 3 and special education. Based upon our review of the literature and our theoretical framework, it is likely that the features of the RTI implementation process are necessary, but that they may not be sufficient to help all children learn to read. Rather, how they are implemented will also impact student outcomes. It is critical to understand more about which system supports for RTI have leverage to maximize the potency of the overall system, but also of tier 3 and special education specifically. Such research may require a mix of sophisticated modeling to account for students being nested in intervention groups, but also some mixed methods and qualitative studies to better unpack the contextual variables that best support RTI.

References

Adams, M. J., Bereiter, C., Brown, A., Campione, J., Carruthers, I., Case, R., et al. (2000). *Open court reading*. Columbus: SRA.

Al Otaiba, S., & Torgesen, J. (2007). Effects from intensive standardized kindergarten and first grade interventions for the prevention of reading difficulties. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *The handbook of response to intervention: The science and practice of assessment and intervention* (pp. 212–222). New York: Springer.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first grade response to intervention (RTI) in reading. *Exceptional Children*, *81* (1) 11–27. doi:10.1177/0014402914532234.

Allor, J. H., & Mathes, P. (2012). *Early interventions in reading: Level K*. Columbus: SRA/McGraw-Hill.

Allor, J. H., Mathes, P., Roberts, J. K., Cheatham, J. P., & Champlin, T. M. (2010a). Comprehensive reading instruction for students with intellectual disabilities: Finding from the first three years of a longitudinal study. *Psychology in the Schools, 47,* 445–466.

Allor, J. H., Mathes, P., Roberts, J. K., Jones, F., & Champlin, T. (2010b). Teaching students with moderate intellectual disabilities to read: An Experimental examination of a comprehensive reading intervention. *Education and Training in Developmental Disabilities, 45,* 3–22.

Allor, J. H., Champlin, T. M., Gifford, D. B., & Mathes, P. G. (2010c). Methods for increasing the intensity of reading instruction for students with intellectual disabilities. *Education and Training in Autism and Developmental Disabilities, 45,* 500–511.

Allor, J. H., Mathes, P., Roberts, K., Cheatham, J. P., & Al Otaiba, S. (2014). Is scientifically-based reading instruction effective for students with Below-Average IQs? *Exceptional Children*, *80*, 289–308. doi:10.1177/0014402914522208.

Berkeley, S., Bender, W. N., Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42,* 85–95.

Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Delzell, L., & Algozzine, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children, 72,* 392–408.

Brown, A. L., Palincsar, A. S., & Purcell, L. (1986). *The school achievement of minority children: New perspectives* (pp. 105–143). Hillsdale: Lawrence Erlbaum Associates, Publishers. (Poor readers: Teach, don't label).

Burns, M. K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37,* 126–136.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102,* 327–340.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., Barquero, L. A., & Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45,* 204–216.

Daggett, W. R. (2012). Achieving reading proficiency for all. http://www.leadered.com/pdf/Reading%20White%20Paper.pdf. Accessed 1 March 2015.

Daly, E. J., III, & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27,* 459–469.

Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26,* 554–574.

Daly, E. J., III, Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8,* 203–218.

Daly, E. J., III, Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 83–94.

Daly, E. J., III, Andersen, M., Gortmaker, V., & Turner, A. (2006). Using experimental analysis to identify reading fluency interventions: Connecting the dots. *The Behavior Analyst, 7*(1), 133–150.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston: Council for Exceptional Children.

Denton, C. A. (2012). Response to Intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities, 45,* 232–243.

Denton, C. A., Fletcher, J. M., Anthony, J. L., & Frances, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities, 39,* 447–466.

Eckert, T. L., Ardoin, S. P., Daly, E. J., III, & Martens, B. K. (2002). Improving oral reading fluency: a brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35,* 271–281.

Engelmann, S., & Bruner, E. C. (1995). *Reading mastery I I/I fast cycle*. Worthington, DC: SRA Macmillan/McGraw-Hill.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3–17.

Fuchs, L. S., & Fuchs, D. (1986). Linking assessment to instructional intervention: An overview. *School Psychology Review, 15,* 318–323.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45,* 195–203.

Fuchs, D., Compton, D. L., Fuchs, L. S., Hamlett, C. L., & Lambert, W. (in press). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities*.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18,* 157–171.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27,* 216–228.

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76,* 301–323.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78,* 263–279.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Harn, B. A., Chard, D. J., Biancarosa, G., & Kame'enui, E. J. (2011). Coordinating instructional supports to accelerate at-risk first-grade readers' performance. *Elementary School Journal, 112,* 332–355.

Hoover, J. J., Eppolito, A., Klingner, J. K., & Baca, L. (2012). Collaborative decision-making in multicultural contexts. In J. B. Crockett, B. S. Billingsley, & M. L. Boscardin (Eds.), *Handbook of leadership and administration for special education* (pp. 191–208). New York: Routledge.

Ihnot, C., Mastoff, J., Gavin, J., & Hendrickson, L. (2001). *Read naturally*. St. Paul: Read Naturally.

Jenkins, J. R., Schiller, E., Blackorby, J., Kalb Thayer, S., & Tilly, W. D. (2013). Response to intervention in reading: architecture and practices. *Learning Disabilities Quarterly, 36*(1), 36–46.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Johnston, J. M., & Pennypacker, H. S. (1993). *Readings for strategies and tactics of behavioral research* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Klingner, J. K., & Harry, B. (2006). The special education referral and decision-making process for english language learners: Child study team meetings and placement conferences. *Teachers College Record, 108,* 2247–2281.

Lemons, C. J., & Fuchs, D. (2010). Modeling response to reading intervention in children with down syndrome: An examination of predictors of differential growth. *Reading Research Quarterly, 45,* 134–168.

Lentz, F. E., Jr. (1988). Effective reading interventions in the regular classroom. In J. L. Graden, J. E. Zins, & M L. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 351–370). Washington, DC: National Association of School Psychologists.

Lonigan, C. J., & Shanahan, T. (2009). *Developing early literacy: Report of the national early literacy panel. executive summary. A scientific synthesis of early literacy development and implications for intervention*. Washington, DC: National Institute for Literacy.

Maslen, B. L., & Maslen, J. R. (1987). *The Bob books*. West Linn: Bob Books Publications.

Mathes, P. G., & Torgesen, J. K. (2005). *Early interventions in reading, Level 1*. Columbus: SRA/McGraw-Hill.

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005).

The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40,* 148–182.

McGuiness, C., McGuiness, D., & McGuiness, G. (1996). Phono-Graphix: A new method for remediating reading difficulties. *Annals of Dyslexia, 46,* 73–96.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice, 24,* 186–195.

Mellard, D., McKnight, M., & Jordan, J. (2010). RtI tier structures and instructional intensity. *Learning Disabilities Research & Practice, 25,* 217–225.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., et al. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659–663.

Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39,* 335–355.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. http://www.nichd.nih.gov/publications/nrp/smallbook.htm. Accessed 1 March 2015.

O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. [RtI]. *Journal of Learning Disabilities, 38,* 532–538.

Orosco, M., & Klingner, J. (2010). One school's implementation of RTI with English language learners: 'Referring into RTI.' *Journal of Learning Disabilities, 43,* 269–288.

Petursdottir, A., McMaster, K., McComas, J. J., Bradfield, T., Braganza, V., Koch-McDonald, J., Rodriguez, R., & Scharf, H. (2009). Brief experimental analysis of early reading interventions. *Journal of School Psychology, 47,* 215–243.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education, 1,* 1–57.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.

Snow, C. E. (2002). *Reading for understanding: Toward an R & D program for comprehension*. Santa Monica: RAND.

Snow, C. E., Burns, S. M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

SRA/McGraw-Hill. (1998). *Phonics mini books*. Columbus: SRA/McGraw-Hill.

SRI International. (2002). Special education elementary longitudinal study (SEELS) [Data tables and reports]. http://www.seels.net/grindex.html. Accessed 1 March 2015.

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: Guilford Press.

Texas Education Agency. (1998). *Texas primary reading inventory (TPRI)*. Austin: Author.

Torgesen, J. K. (2005). *Recent discoveries on remedial interventions for children with dyslexia*. Malden: Blackwell.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading* (pp. 161–188). Mahwah: Lawrence Erlbaum Associates Publishers.

Torgesen, J. K., Wagner, R., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin: PRO-ED.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34,* 33–58.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Vaughn, S., Wanzek, J., Linan-Thompson, S., & Murray, C. S. (2007). Monitoring response to supplemental services for students at risk for reading difficulties: High and low responders. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 234–243). New York: Springer.

Vaughn, S., Linan-Thompson, S., Woodruff, T., Murray, C., Wanzek, J., Scammaca, N., et al. (2008). Effects of professional development on improving at-risk students' performance in reading. In C. R. Greenwood, R. R. Kratochwill, & M. Clements (Eds.), *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 115–142). New York: Guilford Press.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75,* 165–183.

Vellutino, F. R., Scanlon, D. M., Small, S. G., & Fanuele, D. P. (2006). Response to Intervention as a vehicle for distinguishing between reading disabled and non-reading disabled children: Evidence for the role of kindergarten and first grade intervention. *Journal of Learning Disabilities, 39,* 157–169.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36,* 541–561.

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students demonstrating insufficient response to intervention. *Journal of Learning Disabilities, 41,* 126–142.

Wanzek, J., & Vaughn, S. (2010). Tier 3 interventions for students with significant reading problems. *Theory into Practice, 49,* 305–314.

Wanzek, J., & Vaughn, S. (2011). Is a three-tier reading intervention model associated with reduced placement in special education? *Remedial and Special Education, 32,* 167–175.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children, 78,* 89–106.

Werts, M., Lambert, M., & Carpenter, E. (2009). What special education directors say about RTI. *Learning Disability Quarterly, 32,* 245–254.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69,* 848–872.

Wiederholt, J. L., & Bryant, D. R. (2002). *Gray oral reading test* (4th ed.). Austin: PRO-ED.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89,* 420–432.

Witt, J. C., Daly, E. M., & Noell, G. (2000). *Functional assessments: A step by step guide to solving academic and behavior problems*. Longmont: Sopris West.

Zirkel, P. A., & Thomas, L. B. (2010). State laws and guidelines for implementing RTI. *Teaching Exceptional Children, 43*(1), 60–73.

Part VI

Contemporary Implementation Science

Assuring the Response to Intervention Process Has Substance: Assessing and Supporting Intervention Implementation

George H. Noell and Kristin A. Gansle

The conceptual model underlying response to intervention (RTI) as an approach to serve all children and youth in schools has enormous philosophical appeal. Systematically screening so that educators can proactively meet all students' needs makes sense. Similarly, providing intervention supports when they are needed at an intensity that matches students' needs is a rational choice. However, RTI as the foundation of service delivery is like many appealingly rational ideas: The devil is in the details. As the companion chapters in this volume illustrate, the technical and practical choices surrounding doing the right thing are consequential. For example, the decisions surrounding what skills and behaviors to screen, how frequently to screen them, what tools to use to do the screening, and how to make decisions regarding further assessment or intervention are thorny both at a technical and at a practical level (Shapiro et al. 2012). It is obvious that there is a lack of both the depth and the breadth of research that would be helpful to make data-informed decisions regarding these issues (Burns et al. 2010; Gilbert et al. 2012). Similarly and perhaps more severely, the sufficiency of the literature base to inform the subtler and continuous decisions around intervention design, progress monitoring design, program revision, and intervention intensity is similarly uneven.

Although some might read the opening comments as a call for deferral of implementation of RTI to some future date when new research might provide more informed choices, that is not the argument the authors would advance. Sustaining a referral- and diagnosis-based system that has accumulated considerable evidence that it does not serve many children well and in many instances leads to irrational decisions (e.g., wait to fail) is a patently irrational choice. Current students' education cannot be deferred until some future in which there might be more comprehensive research. Deferring change may be more comfortable in the present, but its long-term consequences are likely to be dire. Educators have to make decisions *now* about how to support the children who are at hand *now*. They do not have the option of postponing decisions to a later time at which the choices might be made easier by the emergence of new facts. The best evidence, professional judgment, and action that are available *now* are far more valuable to a student who is struggling today than what might be available in 5 years.

When it is implemented as designed, RTI should be a self-correcting and improving system based on a virtuous circle of information gathering, action, further information gathering, reappraisal, and refinement (Mercier Smith et al. 2009). This virtuous circle is analogous to

G. H. Noell (✉) · K. A. Gansle
Louisiana State University, Baton Rouge, LA, USA
e-mail: gnoell@lsu.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_24

but partially distinct from cycles of advance and continuous improvement evident in the sciences, engineering, and management (Locke and Jain 1995). In all of these cases, initial errors, whether they are the result of faulty evidence (e.g., measurement error) or judgment, *should* be less devastating than in more discrete decision-making systems (e.g., diagnostic systems) because they are simply starting positions. Those starting positions are initial plans based on the evidence and judgment at hand whose utility is continually reevaluated in the context of ongoing action and the emergence of new information. As with the emergence, rejection, and replacement of concepts such as the static heliocentric universe, the fact that the original evaluation of the data was demonstrably wrong did not prevent progress leading to stronger models such as the current inflationary view of the universe (Greene 2011). The critical feature of these systems of continual progress or improvement is the ongoing testing of working hypotheses and actions against new information with consequent revision of assumptions and actions.

The practical and philosophical appeal of RTI is built on bedrock assumptions regarding the integrity of the process that may be more tenable in theory than in practice. If RTI practitioners do not collect relevant data and evaluate screening or progress monitoring data in a consistent manner, the system breaks down. Functionally, there would be no evidential check of current decisions. If the planned interventions are not implemented as designed, then the available data do not provide a check of current working hypotheses regarding student needs (Gresham 1989). Finally, if the process of evaluating current implementation and outcomes is not implemented with substantive rigor, the continuous improvement and timely refinement elements of RTI's raison d'être is undermined. Unfortunately, a consequential database now exists demonstrating that implementation of interventions is likely to be poor in schools unless systematic programming is applied to assure its implementation (Noell et al. in press).

The risks surrounding poor implementation of RTI in education are relatively unique and severe compared to other disciplines. For example, in the sciences, review of the same phenomena by independent laboratories who are seeking to publish their results in peer-reviewed journals makes the long-term survivability of faulty data or analysis very unlikely indeed (Kuhn 1996). Unfortunately, implementation of RTI in schools lacks the transparency and visibility of publication and peer review to drive rigor. Similarly, in engineering and business, the quality of the data analytic process is continually checked by the market and competitive pressures. Organizations that use poor-quality data or implement processes poorly eventually will be driven out of the market. Unfortunately, implementation of RTI lacks a competitive process to drive excellence.

What RTI does have are social environmental supports to drive its implementation. Educators' ethical responsibility and personal motivation to meet students' needs certainly plays an important role. However, this driving force has to compete with a complex array of environmental demands in which educators are legally and ethically responsible for a diverse array of competing demands. The risk is consequential that the interventions or processes underlying RTI will not compete successfully for educator time in these complex environments. Another significant driver of RTI implementation should be the reality that the RTI process can tie into special education entitlement decisions for students. In contexts in which RTI forms the basis for special education entitlement decisions, it becomes a fundamental civil rights protection for students, assuring that they are neither denied services to which they are entitled to achieve a free and appropriate education, and that they are not inappropriately segregated (Jacob et al. 2011; Noell 2008). Obviously, an entitlement process in which students' needs are evaluated based on application of interventions that did not actually happen is vacuous and would not stand up to legal scrutiny. However, the practical reality is that legal scrutiny of educational decisions is both infrequent and delayed from service provision. Unfortunately, delayed, infrequent consequences tend not to be very effective in maintaining behavior.

In sum, RTI is an intellectually appealing model for identifying and meeting student needs in a timely fashion that is congruent with how progress has been achieved through human history in the sciences, economics, and engineering. Like those disciplines, the actual utility of RTI depends on the extent to which relatively complex and ongoing processes are implemented with integrity. Unlike the sciences and economic markets, RTI in education lacks the peer review, independent testing, and market competition that put continual corrective pressures for procedural integrity on human activity. In contrast, the societal supports for implementation of RTI in education are relatively weak and have to compete with powerful immediate environmental demands that educators allocate their limited time to other behaviors. The imbalance between the critical importance of the integrity of the process and competing demands for other behaviors is evident in a relatively extensive and growing literature demonstrating poor implementation of interventions in education without programmed support (Belfiore et al. 2008; Codding et al. 2005; Jones et al. 1997; Noell et al. in press; Ward et al. 1997).

Collectively, the available facts suggest that specifying an RTI model, training educators, and hoping that educators will implement it is patently insufficient. Schools need robust practical implementation support systems. Logically, these systems might draw on the social accountability, peer review, and competitive environmental supports that have proven successful in other disciplines. However, alternative approaches that are better aligned to the educational enterprise are worth considering. The balance of this chapter begins with a brief conceptual overview and is followed by an examination of critical issues in the measurement and support of the implementation of interventions in schools and how implementation may moderate outcomes. The chapter concludes with a discussion of the implications for practice and needs for future research.

What Is Treatment/Procedural Integrity, Fidelity, or Implementation?

As this section's baroque title is intended to suggest, the vocabulary surrounding implementing plans in the applied social sciences is complex and frankly is becoming more complicated. This increasing complexity is discussed below under measurement. A key challenge in understanding the intervention implementation literature is the array of terms that have been used to describe intervention implementation. What is being implemented has been described as treatment, intervention, program, practice, and innovation. The operational definitions of as well as the traditions for the use of these terms vary across literatures. For example, *treatment* is the accepted term in clinical and experimental contexts, whereas *intervention* or *program* is common in the educational literature. The degree or kind of implementation has been described as integrity, fidelity, reliability, adherence, differentiation, and implementation (Henggeler et al. 1997; Peterson et al. 1982; Schulte et al. 2009). No theme other than author preference is readily apparent in the selection of terms.

Treatment integrity (TI) appears to be the most frequently used combination of terms, but this terminology is problematic due to its experimental roots. TI is to the extent to which an independent variable (IV) is implemented as designed in an experiment (Peterson et al. 1982). The parallel between implementing an IV in an experiment and implementing an intervention in a school is obvious. However, in reality, implementing interventions in schools and implementing IVs in research are very different. In research, typically, there are researchers whose only reason to be in the context and whose sole responsibility is to implement the research protocol. If they do not implement the research protocol, there is simply no reason for them to be there. In contrast, educators, peers, and parents have a great many competing responsibilities and intervention implementation is normally an added responsibility. It has been argued previously that it is normally harder to get interventions implemented than it is to identify an effective intervention and the relevant data at least indirectly support this contention (Foxx 1996; Noell et al. 2005).

This chapter follows the conventions suggested by Noell (2008) in discussing the implementation of practices in education. *TI* is used to describe the implementation of the independent variable in an experimental study, but not to describe application of interventions in practice. It is worth noting at the outset that the consequential literature regarding the assessment and failure to assess TI in the research literature is irrelevant to practice (see Barnett et al. in press for a recent review). *Intervention plan implementation* (IPI) describes the degree to which an intervention is implemented as designed. Intervention was adopted over treatment because that is the most common term used in schools.

Definition and Measurement of Intervention Plan Implementation

IPI and RTI-PI are complex, multidimensional human behaviors and as a result, operationally defining and measuring them are significant challenges. These challenges are magnified by the array of specific RTI processes and the diversity of interventions that can be deployed within them. Potentially, IPI can be assessed along a variety of dimensions such as the treatment agents understanding the rationale for the intervention, understanding the steps of the procedure, and being able to enact the procedure (Schulte et al. 2009). To some degree, all are desirable training outcomes that are distinct from enacting the intervention or process in the natural environment at the time it is needed. The definition and measurement of the intervention might be accomplished at a highly molar level such as providing more attention for pro-social behavior, or at an extremely molecular level based on hundreds of discrete motor and vocal responses by the treatment agent.

The existing research has consistently adopted an intermediate level of measurement, focusing on the occurrence or nonoccurrence of a moderate number of discrete, observable steps of intervention (e.g., DiGennaro et al. 2007; Sanetti et al. 2012). This intermediate level of definition of IPI has emerged partially due to practical reasons and because it has been successful. At a practical level, it can be measured reliably, avoiding the difficult discriminations that global definitions require, and avoiding the tremendous burdens imposed by the necessity of measuring extremely microanalytic pieces of intervention. It has been successful in providing a level of data analysis that has met the major needs of applied researchers and practitioners. First, this intermediate level of IPI measurement has been shown to vary across individuals and over time (e.g., DiGennaro et al. 2005; Mortenson and Witt 1998; Noell et al. 1997). Second, IPI defined at an intermediate procedural level has been shown to be sensitive to environmental conditions and experimentally controllable (Fox et al. 2011; Rodriguez et al. 2009). Third, IPI measured as number or percentage of procedural steps completed has been demonstrated to be related to intervention outcome (Noell et al. 2000; Noell et al. 2005; Sanetti et al. 2012).

Although the accumulated evidence demonstrates that the procedural level definition of IPI is useful, a number of challenges remain. For example, researchers and practitioners generally recognize that all steps of an intervention may not be equally important and that the importance of various steps may change over the course of treatment (Noell and Gansle 2006; Schulte et al. 2009). For example, the prompting component of constant time delayed instruction may be particularly important during initial acquisition, but may diminish relative to feedback and reinforcement components as the student progresses. Although it is clear that intervention components vary in their difficulty of implementation and importance to treatment outcome, there is a lack of data for applying empirically derived weighting in most cases (Noell and Gansle 2006; Schulte et al. 2009). In the face of this uncertainty, researchers have generally applied equal weights to all steps. This approach is based in part on findings supporting the robustness of improper linear models with unit weights (Dawes 1979). Although this may be the most robust approach in the face of uncertainty, detailed research examining specific interventions has demonstrated that all steps are not equally important for all students and interventions (Leon et al. in press).

As the scientific examination of IPI has emerged as an issue for the applied social sciences over the past two decades, researchers have identified the multidimensional nature of IPI as a challenge for understanding and controlling implementation. In simple terms, IPI can vary in multiple ways. For example, a tutor may simply omit praise for correct responding, or may provide praise with a sour, sarcastic tone that may change the putative function of praise. This might be described as a quantitative versus a qualitative distinction. Clearly, IPI varies in the quality of implementation, the quantity of implementation, intervention efficacy, and the contextual adaptation of interventions (Power et al. 2005). A meta-analysis of the effect of implementation integrity on primary and secondary prevention program outcomes described treatment implementation as including adherence, exposure, delivery quality, program differentiation, and responsiveness to treatment (Dane and Schneider 1998).

Schulte et al. (2009) described TI as consisting of three primary constructs: *treatment delivery, treatment receipt,* and *treatment enactment*. The *treatment delivery* construct appears to be most aligned to IPI. It includes three sub-constructs. First, *adherence* is the total number of or proportion of the total of intended treatment elements delivered. Second, *exposure* is the volume or amount of treatment that is provided by the treatment agent. This would include data like the number of sessions implemented or how often the treatment was delivered. Third, *quality* is the extent to which the treatment delivered would be considered ideal or mastery practice. Fourth, *program differentiation* is the extent to which the treatment that was delivered represented elements of the intervention that was planned originally and not of some unplanned treatment. *Treatment receipt* is described as including exposure to the treatment, comprehension or understanding of the treatment, and responsiveness to intervention. *Treatment enactment* is described as the degree to which the client is able to demonstrate the skills targeted by the intervention. This might also be described as treatment outcome. Other authors have emphasized the ability of those delivering the intervention to do so competently, the dosage of the intervention, the agents' procedural adherence, and participant responsiveness (Jones et al. 2008).

Despite the logical and conceptual appeal of these more complex, multifaceted definitions of IPI, they have not yet led to data demonstrating that they meet the same professional and scientific quality thresholds as procedural adherence. Specifically, there is a lack of data demonstrating that these distinct constructs can be reliably and practically measured, that they vary across individuals and time, that they are responsive to environmental events (i.e., are controllable), and that they are related to outcomes. Additionally, there is a lack of data demonstrating that including these more elaborate constructs leads to better control over intervention outcomes as compared to the more parsimonious and well-researched procedural adherence. Although these multifaceted integrity constructs have obvious rational appeal, at present there is a lack of evidence that they contribute to better intervention outcomes.

If the extant research points to the conclusion that procedural adherence (i.e., step or completion based) measures of IPI are useful, that does not clarify how to measure completion of those steps. The methodological challenges and options in measuring IPI are the same as the measurement of other behaviors in the natural environment. Direct observation provides the lowest inference measure, but the resource demands for obtaining an adequate sample are daunting (Hintze and Matthews 2004) and it may engender reactivity. Permanent product strategies can capture data across time efficiently and have been successful in applied research (e.g., Resetar et al. 2006), but are limited in the types of intervention steps they can capture and are vulnerable to additional sources of measurement error (Haynes and O'Brien 2000). Self-report is extremely resource efficient, but is highly susceptible to problems of bias due to issues such as memory and social desirability. Based on findings suggesting severe upward bias in self-reports, they have not generally been recommended (Noell et al. 2005; Sanetti and Fallon 2011; Sanetti and Collier-Meek in press; Wickstrom et al. 1998). Only recently have data become available comparing measurement of IPI by observation and permanent products. Sanetti and Collier-Meek (in press) found moderate agreement, but the source of the divergence was unclear since there were differences in both measurement method and observational period. The permanent product method captured a much larger total intervention period than direct observation.

In summary, measuring IPI through procedural adherence or step completion has been demonstrated to be practical, responsive to intervention, related to intervention outcome, to vary across individuals, and to vary over time. Measurement of IPI by both direct observation and permanent products has been deployed successfully in applied research, while self-reports have not been successful (Noell et al. 2005; Sanetti and Fallon 2011; Sanetti and Collier-Meek in press; Wickstrom et al. 1998). More recent work has raised the issue that IPI is a complex, multidimensional construct and that IPI can go awry in ways other than simple adherence to implementing the steps of the plan (Power et al. 2005; Schulte et al. 2009), but data have not yet emerged demonstrating which of these additional dimensions are scientifically or practically useful.

Supporting Intervention Plan Implementation

Modern intervention and prevention activities in schools such as RTI that are based on intervention in the natural environment at the time the problems emerge with increasing intensity matched to children and youth's needs places considerable demands on teachers, peers, and parents, as well as educational specialists to implement interventions in addition to completing other routine ongoing activities (Noell 2008). This shift in the design of educational services is predicated in part consistent with findings that early intervention and intervention in the natural context are more effective than alternative services (DuPaul and Eckert 1997; Swanson and Hoskyn 1998; Weiss and Weisz 1995). Additionally, RTI is based on a desire to provide a rational system of support in which all students can be supported at the relevant time and in the most inclusive environment practical. Although the broad empirical basis for RTI is sound and the rational appeal is compelling, the utility of RTI is predicated on the implementation of interventions as planned. Unfortunately, a consequential and growing literature demonstrates that implementation of interventions in education is frequently poor and deteriorates over time (Noell et al. in press).

The initial theoretical work surrounding consultation and intervention in schools argued that IPI would follow a rational course. Various authors theorized that the acceptability of the intervention, understanding of the intervention, and intervention complexity would drive initial use with experienced efficacy further supporting or undermining use (Eckert and Hintze 2000; Gresham 1989; Martens et al. 1985). However, subsequent research has not supported the importance of acceptability, understanding, or complexity. High acceptability has not been found sufficient to assure IPI (Noell et al. 2005). Poor implementation has been found for both very simple interventions and relatively complex interventions (Noell et al. 2000; Witt et al. 1997). Finally, poor implementation has emerged following pre-intervention training in which the intervention agent has demonstrated mastery of the intervention (Taylor and Miller 1997; Noell et al. 2000; Witt et al. 1997).

The existing literature demonstrates that in the absence of procedures specifically designed to support IPI, it is typically poor and deteriorates over time (Gilbertson et al. 2007; Noell et al. in press). Clearly, there are forces at work beyond treatment agent intention and perception that drive behavior change. Schools are complex behavioral systems in which the behavioral regularities of educators are maintained by an existing ecology whose antecedents and consequences have selected for and maintain existing behaviors (Martens and Witt 1988). Educators who implement interventions are under many simultaneous competing demands for their time and attention and over the long term should be expected to allocate their behavior in a way that aligns with the contingencies of the natural environment. It is important to recognize however that when provided sufficient supports adults can be successful in achieving very challenging behavior change (e.g., Webb et al. 2010). The critical reality is that schools have not typically provided systematic support for intervention implementation. For example, educators who are clearly accountable to parents and/or to a principal and provided appropriate follow-up may find this social contingency sufficient to support implementation of an intervention (Noell et al. 2000). The research surrounding factors that may support IPI can be organized around those procedures provided prior to independent implementation and those that occur as follow-up.

Antecedent Procedures to Improve Implementation

Several authors have hypothesized that thorough training might be sufficient to sustain IPI. A study comparing didactic instruction, modeling, and rehearsal with feedback found that the two more direct training procedures yielded higher implementation of an intervention for a tic than did didactic instruction (Sterling-Turner et al. 2001). However, the relevance of this study to interventions in schools is unclear as the research participants were undergraduate students who were implementing the intervention in a laboratory analogue context. In contrast to the analogue study above, studies in schools have been less encouraging regarding the importance of training. Taylor and Miller (1997) used in vivo didactic training to instruct teachers to 100 % accuracy in the use of a time-out procedure. This did not lead to sufficient IPI following training. A number of studies demonstrated poor and deteriorating implementation following in vivo training to 100 % accurate implementation (e.g., Gilbertson et al. 2007; Noell et al. 2000; Witt et al. 1997). Finally, a recent meta-analysis of 29 single subject experimental analyses of IPI did not find significant effects of pre-baseline training including courses in behavior analysis or behavioral principles, discussion or didactic instruction regarding the intervention, modeling, prompting and correcting, practice, and role play did not yield significant effects on baseline levels of implementation (Noell et al. in press).

The description above is not intended to indicate that training cannot have beneficial effects or that it is unnecessary. What the data demonstrate is that among the published studies to date, training does not seem to play a role in sustaining implementation after intervention agents have acquired the necessary skills. These data simply illustrate again that the processes that are necessary to establish new behaviors frequently differ from those that are useful in sustaining and maintaining behavior change. In the case of IPI, this maintenance is commonly the critical challenge.

A common feature of the consultation literature has been a fundamental assumption that the nature of the relationship between the person implementing the intervention (consultee) and the person helping to develop that intervention (consultant) should influence IPI (Garbacz et al. 2008). At the most extreme ends of the continuum, the consultant can be perceived by the consultee as a coequal collaborator or as the director of the intervention development and implementation process (Gutkin 1999). In the only study of which the authors are aware that compared a prescriptive and collaborative approach to intervention development and implementation (all else being the same), teachers collected about half of the required data and implemented programmed intervention consequences less than one time in 20 occasions across conditions, with no significant difference between collaborative and directive conditions (Wickstrom et al. 1998). Taking an alternative approach to social influence, a subsequent study examined the effect of a social influence attempt by consultants emphasizing the teacher's commitment to implement the intervention and the potential adverse consequences of failing to do so (Noell et al. 2005). Although there was a favorable trend in the data, the social influence and commitment emphasis procedure was not associated with higher IPI or improved student outcome. Although it is entirely rational to assume that the quality of the relationship between consultants and intervention agents will influence IPI, at present there is a lack of scientific evidence about the dimensions that matter, how to manipulate those dimensions, or the magnitude of positive results expected to be obtained. More globally, at present there is a lack of data to demonstrate how to predict or control IPI based on procedures that occur prior to implementation.

Follow-Up Procedures to Improve Implementation

Across a considerable number of single subject studies and one randomized field trial IPI in the absence of systematic efforts to support implementation have been consistent, IPI has been low and has frequently decreased with time (Noell et al. in press; Noell et al. 2005). While IPI under conditions that might have been considered state-of-the-art practice 20 years ago are discouraging, procedures developed over the past 15 years through field research have been successful in sustaining implementation. Performance feedback (PFB) regarding implementation has been the most widely successful intervention to support IPI in the research literature (Codding et al. 2005; DiGennaro et al. 2005, 2007). PFB in the context of IPI has consisted of providing summary information to a treatment agent regarding their implementation (Alvero et al. 2001; Codding et al. 2005). Generally, reviews in both education and business have found that PFB is an effective approach to change behavior (Alvero et al. 2001; Balcazar et al. 1985; Casey and McWilliam 2011; Noell et al. in press).

Successful applications of PFB have varied widely. PFB has been effective with individual teachers, with groups, or teams (e.g., Belfiore et al. 2008; Burns et al. 2008; Pellecchia et al. 2011). Coaching during IPI, or immediate PFB, has demonstrated positive outcomes on implementation behavior (Fox et al. 2011; Rodriguez et al. 2009). Early studies demonstrated that daily feedback regarding implementation was effective and subsequent studies have demonstrated procedures for fading or implementing feedback at intervals up to 2 weeks (Codding et al. 2005; Noell et al. 1997; Witt et al. 1997). Research has also demonstrated that follow-up meetings without review of data are not consistently effective (Noell et al. 2000). Graphing implementation data can also improve the effect of the PFB on IPI (Duhon et al. 2009; Noell et al. 2002a; Sanetti et al. 2007).

Recent research has demonstrated the efficacy of a number of procedural variations on PFB. Response-contingent PFB has been found effective in maintaining implementation (DiGennaro et al. 2007, 2005; Gilbertson et al. 2007). Under response-contingent PFB, the follow-up session is canceled if the intervention agent implements the intervention completely. Response-contingent PFB has been conceived of as a negative reinforcement contingency in which implementing the intervention allows the intervention agent to escape having the meeting. Response-contingent PFB has also been successfully combined with feedback fading (Gilbertson et al. 2007). Additionally, some, but not all, of the successful applications of response-contingent PFB have included directed rehearsal of the omitted intervention steps during the feedback meeting (DiGennaro et al. 2007, 2005; Ward et al. 1998). However, results across studies do not indicate a consequential improvement in PFB efficacy with the inclusion of directed rehearsal (Noell et al. in press).

Two recent studies have demonstrated successful approaches to sustaining IPI of behavior analytic interventions provided to students with autism that are not based on PFB. Petscher and Bailey's (2006) study used a pager to prompt teachers to perform specific intervention steps, after which the teachers filled out a self-monitoring form about their accuracy of implementation. Belfiore et al. (2008) videotaped teachers providing discrete trial instruction to children with autism. Following the teaching session, teachers watched the tape of their session and scored the session for the accuracy of implementation. These studies represent the emergence of increasing options in how to support IPI. A recent meta-analysis of the single subject literature examining procedures to support IPI provides a more detailed review of these studies (Noell et al. in press).

The Link Between Implementation and Outcomes

When intervention plans are well conceived and outcomes are measured in relevant ways, IPI predicts outcomes. Compelling evidence is available in myriad journals in education, psychology, and behavior analysis for at least a century demonstrating an absence of behavior change in baseline or control conditions and improvement in human functioning with the onset of intervention. This demonstration is at the heart of all experiments in the applied social sciences (Sidman 1960). The more complex and nuanced questions surrounding the degree to which IPI is important, how variation in IPI over time matters, and how variation in specific elements of IPI matter are much less well understood.

The literature linking variation in IPI and student outcomes is small, but emerging. Studies have examined the impact of IPI across a number of concerns including compliance (Leon et al. in press; Wilder et al. 2006), disruptive behavior (Gansle and McMahon 1997), anxiety disorders (Vermilyea et al. 1984), classwide peer tutoring (Greenwood et al. 1992), social skills training (McEvoy et al. 1990), constant time delay instruction (Holcombe et al. 1994), differential reinforcement (St. Peter Pipkin et al. 2010; Vollmer et al. 1999), strategy instruction (Noell et al. 2002b), and multisystemic therapy for juvenile offenders (Henggeler et al. 1997), among a few others. The diversity of findings, methods, and interventions makes synthesizing results across studies challenging. A critical limitation of this literature, as contrasted with the literature examining supporting implementation, is that no substantive sustained line of inquiry has emerged in which variations in IPI have been tested in a systematic way across studies or investigations to allow for the accumulation of a detailed understanding of the relationship between IPI and outcomes (Noell in press).

Even in the context of such varied studies, some general patterns have emerged in the literature. The relationship between IPI and outcomes appears to be probabilistic in that some decrements in IPI may reduce outcomes and others may result in ineffective treatment, but substantive variability exists across individuals (Henggeler et al. 1997; Noell et al. 2005; St. Peter Pipkin et al. 2010; Wilder et al. 2006; Leon et al. in press). Although the relationship between intervention response and IPI might look continuous and consistent when examined in aggregate, this hides underlying individual variability in response. It is also the case that some degradation in IPI for some interventions and individuals has not resulted in any loss in treatment effectiveness (Gansle and McMahon 1997; Holcombe et al. 1994; Vollmer et al. 1999; Wilder et al. 2006).

Some more nuanced and systematic research is available specifically examining differential reinforcement procedures (Leon et al. in press; St. Peter Pipkin et al. 2010, Vollmer et al. 1999). Consistent with the studies above, these studies found complete IPI effective and also found individual differences in response to decreased IPI. They found that the order in which decreased IPI is experienced matters: Prior exposure to high IPI resulted in poorer response to subsequent lower IPI (Leon et al.). Intermediate IPI was more effective following baseline than following 100% IPI. Leon et al. also found that errors of commission were more damaging than errors of omission in the differential reinforcement of compliance.

It is clear that what is sorely needed at this point is systematic, sustained research examining a few widely used, well-understood interventions that clarify what the critical dimensions of IPI are and what the levels of implementation are at which results are normally robust. This would provide a more firm basis of literature from which to begin to generalize to the diverse interventions deployed in schools, communities, and clinics. The utility of this line of research will be greatly increased if it also attends to the frequently noted individual differences in response to decreased IPI and the emerging evidence that sequence effects may be important in response to IPI (Leon et al. in press). This type of research could quantify typical expected intervention response (e.g., effect size), the most critical dimensions of implementation, and the implementation levels at which the interventions would be expected to be robustly effective. Additionally, this type of research could clarify for whom precise IPI is most critical and the relative importance of high IPI at the beginning of the intervention process versus high IPI as progress is made. Is it more critical to start with precise IPI at the outset or is it more important to sustain IPI level as intervention progresses?

The existing literature is quite clear in some regards and disappointingly vague in others. When the types of interventions that are characteristic of the research literature are implemented with complete or nearly complete accuracy, positive effects are consistent and consequential. As IPI decreases, mean intervention outcomes across groups decrease (Noell et al. 2005), but these mean differences mask consequential underlying variation in results in which some individuals respond favorably even at decreased IPI and others do not (Henggeler et al. 1997; St. Peter Pipkin et al. 2010; Wilder et al. 2006). In essence, decreasing accuracy of implementation appears to act as a stressor on the intervention with decreasing IPI increasing the risk of failure. This may be somewhat analogous to weight or load on a bridge. As the load on the bridge is increased (decreasing IPI) the exact point of failure is unknown, but the risk of failure continuously increases to the point of ultimate certainty.

Implications for Practice, Research Needs, and Summary

RTI is an appealing approach to meeting the needs of all students in schools. The rational choice to provide supports at the time they are needed, in the relevant context, and at a level of intensity that is matched to students' needs is difficult to argue against. However, this enormously appealing rational framework is built upon foundational bedrock that exposes consequential vulnerability. If the interventions that are the core of RTI are not implemented as designed, the entire edifice will deservedly collapse in time. Unfortunately and alarmingly for RTI proponents, a substantial literature base has emerged over the past 15 years demonstrating that this threat to the foundation of RTI is quite real and distressingly common (Noell et al. in press). In a more hopeful vein, the same literature that has demonstrated how common poor and deteriorating intervention implementation is has provided a way forward by demonstrating procedures that sustain implementation.

One of the important challenges facing practitioners and researchers who are focused on how to support meaningful implementation is how to define and measure IPI. Interventions are complex behaviors that can be measured in many ways, along multiple dimensions, and at widely varying levels of granularity. A substantial literature exists demonstrating the scientific and practical utility of measurement at intermediate levels of detail (e.g., intervention steps completed, Fox et al. 2011; Rodriguez et al. 2009). Measurement at this level of detail has been demonstrated to vary across individuals, across time, to be related to student outcome, and to be responsive to environmental supports (Noell et al. 2005). More recently, scholars have argued for the importance of more complex multidimensional IPI constructs; however, to date there is a lack of evidence that these more complex constructs are practically necessary or would be more scientifically successful (Dane and Schneider 1998; Power et al. 2005; Schulte et al. 2009). The literature on measurement methods has supported the utility of direct observations and permanent product measures, but has been discouraging regarding the utility of self-report measures (Noell et al. 2005; Sanetti and Collier-Meek in press).

The implications of the literature for practice are extremely direct. IPI needs to be directly assessed by relatively direct measures (observation or permanent products) as a routine part of practice. Failure to assess implementation will likely yield poor and deteriorating implementation resulting in poor student outcomes and a vacuous RTI process. The implications of the extant literature for future research are similarly obvious. Research is needed examining the scientific and practical utility of the more complex IPI constructs that have been proposed. Research is also needed clarifying the relationship between direct observation and permanent product measures, the conditions under which each is most useful, and the utility of composite measures that blend assessment strategies.

The findings supporting IPI are also quite clear. At present, an effective antecedent strategy for sustaining implementation is not known, but follow-up providing anyone a variety of PFB procedures has been demonstrated to maintain IPI (Noell et al. in press). PFB has been demonstrated effective across diverse intervention agents, interventions, and referral concerns. Positive results have been obtained with sessions as frequent as daily and as infrequent as once per 2 weeks (Codding et al. 2005; DiGennaro et al. 2007). PFB has been more successful when it has included graphs and has remained successful when adjunctive procedures such as contingent meeting cancellation and direct rehearsal of omitted steps have been added (DiGennaro et al. 2005; Duhon et al. 2009; Noell et al. 2002a).

The implications of the literature for practice are extremely direct. Follow-up support that includes graphed IPI performance review with another professional should be a routine part of practice. The research needs in this domain are similarly clear. The profession needs additional research exploring effective and cost-efficient procedures that support implementation and provide practitioners a wider variety of evidence-based options. Additionally, it would be an invaluable contribution to the profession if an antecedent intervention were identified that would substantively support implementation. However, the absence of positive findings for antecedent intervention may be more reflective of the many demands on treatment agents in natural environments than a failing of the research literature.

The extant literature clearly demonstrates that the level of IPI is related to intervention outcomes with poor implementation increasing the risk of intervention failure (St. Peter Pipkin et al. 2010; Vollmer et al. 1999). Unfortunately, the specific value of IPI needed to demonstrate positive outcomes is not clear. Across studies, the data are encouraging at 80 or 90 % implementation, but positive results are sometimes obtained at lower values. These approximate guidelines do not address the issue of the varied importance of specific intervention steps or kinds of mistakes (Leon et al. in press). The current implications for practice appear to argue for setting relatively high goals for implementation. Rational and data-based arguments can be made for values in the 80 or 90 % range. The most important consideration may be comparing obtained results to current implementation and making decisions to adjust intervention when they appear needed. For example, if current implementation is 70 %, but RTI is excellent, no further action may be needed. In contrast, if RTI is weak, working to further improve implementation would appear advised. Clearly, a great deal of research is needed experimentally manipulating IPI in varied ways and examining its impact on outcomes.

In summary, IPI is a cornerstone of RTI's promise to meet student needs. It is known that in the absence of systematic follow-up IPI is commonly poor and deteriorates. Follow-up procedures that include review of objective implementation data with another person have been found to be sufficient to sustain higher levels of implementation, and increased IPI has been found to be predictive of improved outcomes. Assessment of implementation and follow-up for implementation should be a routine part of RTI practice.

References

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings. *Journal of Organizational Behavior Management, 21,* 3–29. doi:10.1300/J075v21n01_02.

Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7,* 65–89. doi:10.1300/J075v07n03_05.

Barnett, D., Hawkins, R., Wahl, E., Shier, A., Denune, H., McCoy, D., & Kimener, L. (in press). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education*.

Belfiore, P. J., Fritts, K. M., & Herman, B. C. (2008). The role of procedural integrity: Using self-monitoring to enhance Discrete Trial Instruction (DTI). *Focus on Autism and Other Developmental Disabilities, 23,* 95–102. doi:10.1177/1088357607311445.

Burns, M. K., Peters, R., & Noell, G. H. (2008). Using performance feedback to enhance implementation fidelity of the problem-solving team process. *Journal of School Psychology, 46,* 537–550. doi:10.1016/j.jsp.2008.04.001.

Burns, M. K., Scholin, S. E., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment, 28,* 102–114.

Casey, A. M., & McWilliam, R. A. (2011). The impact of checklist-based training on teachers' use of the zone defense schedule. *Journal of Applied Behavior Analysis, 44,* 397–401. doi:10.1901/jaba.2011.44-397.

Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis, 38,* 205–219. doi:10.1901/jaba.2005.98-04.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23–45. doi:10.1016/S0272-7358(97)00043-3.

Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review, 34,* 220–231.

DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40,* 447–461. doi:10.1901/jaba.2007.40-447.

Duhon, G. J., Mesmer, E. M., Gregerson, L., & Witt, J. C. (2009). Effects of public feedback during RTI team meetings on teacher implementation integrity and student academic performance. *Journal of School Psychology, 47,* 19–37. doi:10.1016/j.jsp.2008.09.002.

DuPaul, G. J., & Eckert, T. L. (1997). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis. *School Psychology Review, 26,* 5–27.

Eckert, T. L., & Hintze, J. M. (2000). Behavioral conceptions and applications of acceptability: Issues related to service delivery and research methodology. *School Psychology Quarterly, 15,* 123–148. doi:10.1037/h0088782.

Fox, L., Hemmeter, M., Snyder, P., Binder, D. P., & Clarke, S. (2011). Coaching early childhood special educators to implement a comprehensive model for promoting young children's social competence. *Topics in Early Childhood Special Education, 31,* 178–192. doi:10.1177/0271121411404440.

Foxx, R. M. (1996). Twenty years of applied behavior analysis in treating the most severe problem behavior: Lessons learned. *The Behavior Analyst, 19,* 225–235.

Gansle, K. A., & McMahon, C. M. (1997). Component integrity of teacher intervention management behavior using a student self-monitoring treatment: An experimental analysis. *Journal of Behavioral Education, 7,* 405–419. doi:10.1023/A:1022851117439.

Garbacz, S., Woods, K. E., Swanger-Gagné, M. S., Taylor, A. M., Black, K. A., & Sheridan, S. M. (2008). The effectiveness of a partnership-centered approach in conjoint behavioral consultation. *School Psychology Quarterly, 23,* 313–326. doi:10.1037/1045-3830.23.3.313.

Gilbert, J. K., Compton, D. L., Fuchs, D., & Fuchs, L. S. (2012). Early screening for risk of reading disabilities: Recommendations for a four-step screening system. *Assessment for Effective Intervention, 38,* 6–14. doi:10.1177/1534508412451491.

Gilbertson, D., Witt, J. C., Singletary, L. L., & VanDerHeyden, A. (2007). Supporting teacher use of interventions: Effects of response dependent performance feedback on teacher implementation of a math intervention. *Journal of Behavioral Education, 16,* 311–326. doi:10.1007/s10864-007-9043-0.

Greene, B. (2011). *The hidden reality: Parallel universes and the deep laws of the cosmos*. New York: Alfred A. Knopf.

Greenwood, C. R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The classwide peer tutoring program: Implementation factors moderating students' achievement. *Journal of Applied Behavior Analysis, 25,* 101–116. doi:10.1901/jaba.1992.25-101.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18,* 37–50.

Gutkin, T. B. (1999). Collaborative versus directive/prescriptive/expert school-based consultation: Reviewing and resolving a false dichotomy. *Journal of School Psychology, 37,* 161–190. doi:10.1016/S0022-4405(99)00003-5.

Haynes, S. N., & O'Brien, W. (2000). *Principles and practice of behavioral assessment*. Dordrecht: Kluwer Academic Publishers.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65,* 821–833. doi:10.1037/0022-006X.65.5.821.

Hintze, J. M., & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review, 33,* 258–270.

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4,* 49–73.

Jacob, S., Decker, D. M., & Hartshorne, T. S. (2011). *Ethics and law for school psychologists* (6th ed.). New York: Wiley.

Jones, K. M., Wickstrom, K. F., & Friman, P. C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12,* 316–326.

Jones, H. A., Clarke, A. T., & Power, T. J. (2008). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. *InBalance, Newsletter of Division 53* (*Clinical Child and Adolescent Psychology*) *of the American Psychological Association, 23*(1), 4–5.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: The University of Chicago Press.

Leon, Y., Wilder, D. A., Majdalany, L., Myers, K., & Saini, V. (in press). Errors of omission and commission during alternative reinforcement of compliance: The effects of varying levels of treatment integrity. *Journal of Behavioral Education*. doi:10.1007/s10864-013-9181-5.

Locke, E. A., & Jain, V. K. (1995). Organizational learning and continuous improvement. *International Journal of Organizational Analysis, 3,* 45–68. doi:10.1108/eb028823.

Martens, B. K., & Witt, J. C. (1988). On the ecological validity of behavior modification. In J. C. Witt, S. N. Elliott, & F. M. Gresham (Eds.), *Handbook of behavior therapy in education* (pp. 325–342). New York: Plenum Press.

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice, 16,* 191–198. doi:10.1037/0735-7028.16.2.191.

McEvoy, M. A., Shores, R. E., Wehby, J. H., Johnson, S. M., & Fox, J. J. (1990). Special education teachers' implementation of procedures to promote social interaction among children in integrated settings. *Education and Training in Mental Retardation, 25,* 267–276.

Mercier Smith, J. L., Fien, H., Basaraba, D., & Travers, P. (2009). Planning, evaluating and improving tiers of support in beginning reading. *Teaching Exceptional Children, 41*(5), 16–22.

Mortenson, B. P., & Witt, J. C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology Review, 27,* 613–627.

Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 323–342). Mahwah: Lawrence Erlbaum Associates, Inc.

Noell, G. H. (in press). Implementation of treatments in research and practice: Commentary on the special issue. *Journal of Behavioral Education*.

Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention, 32,* 32–39. doi: 10.1177/15345084060320010501.

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. L. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12,* 77–88. doi:10.1037/h0088949.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Rainer, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33,* 271–284. doi:10.1901/jaba.2000.33-271.

Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002a). Consultation, follow-up, and behavior management intervention implementation in general education. *School Psychology Review, 31,* 217–234.

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002b). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11,* 51–67. doi:10.1023/A:1014385321849.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Resetar, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34,* 87–106.

Noell, G. H., Gansle, K. A., Mevers, J. L., Knox, R. M., Mintz, J. C., & Dahir, A. (in press). Improving treatment plan implementation in schools: A meta-analysis of single subject design studies. *Journal of Behavioral Education*. doi:10.1007/s10864-013-9177-1.

Pellecchia, M., Connell, J. E., Eisenhart, D., Kane, M., Schoener, C., Turkel, K., Riley, M., & Mandell, D. S. (2011). We're all in this together now: Group performance feedback to increase classroom team data collection. *Journal of School Psychology, 49,* 411–431.

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15,* 477–492. doi:10.1901/jaba.1982.15-477.

Petscher, E. S., & Bailey, J. S. (2006). Effects of training, prompting, and self-monitoring on staff behavior in a classroom for students with disabilities. *Journal of Applied Behavior Analysis, 39,* 215–226. doi:10.1901/jaba.2006.02-05.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42,* 495–507. doi:10.1002/pits.20087.

Resetar, J. L., Noell, G. H., & Pellegrin, A. L. (2006). Teaching parents to use research-supported systematic strategies to tutor their children in reading. *School Psychology Quarterly, 21,* 241–261. doi:10.1521/scpq.2006.21.3.241.

Rodriguez, B., Loman, S. L., & Horner, R. H. (2009). A preliminary analysis of the effects of coaching feedback on teacher implementation fidelity of first step to success. *Behavior Analysis in Practice, 2*(2), 11–21.

Sanetti, L., & Collier-Meek, M. A. (in press). Increasing the rigor of procedural fidelity assessment: An empirical comparison of direct observation and permanent product review methods. *Journal of Behavioral Education*. doi:10.1007/s10864-013-9179-z.

Sanetti, L. M. H., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational & Psychological Consultation, 21,* 209–232. doi:10.1080/10474412.2011.595163.

Sanetti, L. M. H., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public elementary school. *Behavior Modification, 31,* 454–465. doi:10.1177/0145445506297583.

Sanetti, L. M. H., Dobey, L. M., & Gritter, K. L. (2012). Treatment integrity of interventions with children in the *Journal of Positive Behavior Interventions* from 1999 to 2009. *Journal of Positive Behavior Interventions, 14,* 29–46. doi:10.1177/1098300711405853.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38,* 460–475.

Shapiro, E. S., Hilt-Panahon, A., Gischlar, K. L., Semeniak, K., Leichman, E., & Bowles, S. (2012). An analysis of consistency between team decisions and reading assessment data within an RTI model. *Remedial and Special Education, 33,* 335–347. doi:10.1177/0741932510397763.

Sidman, M. (1960). *Tactics of scientific research*. Oxford England: Basic Books.

St. Peter Pipkin, C. C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis, 43,* 47–70. doi:10.1901/jaba.2010.43-47.

Sterling-Turner, H. E., Watson, T., Wildmon, M., Watkins, C., & Little, E. (2001). Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly, 16,* 56–67. doi:10.1521/scpq.16.1.56.19157.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68,* 277–321.

Taylor, J., & Miller, M. (1997). When timeout works some of the time: The importance of treatment integrity and functional assessment. *School Psychology Quarterly, 12,* 4–22. doi:10.1037/h0088943.

Vermilyea, B. B., Barlow, D. H., & O'Brien, G. T. (1984). The importance of assessing treatment integrity: An example in the anxiety disorders. *Journal of Behavioral Assessment, 6,* 1–11. doi:10.1007/BF01321456.

Vollmer, T. R., Roane, H. S., Ringdahl, J. E., & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis, 32,* 9–23. doi:10.1901/jaba.1999.32-9.

Ward, P., Smith, S., & Makasci, K. (1997). Teacher training: Effects of directed rehearsal on the teaching skills of physical education majors. *Journal of Behavioral Education, 7,* 505–517. doi:10.1023/A:1022863520165.

Ward, P., Johnson, M., & Konukman, F. (1998). Directed rehearsal and preservice teachers' performance of instructional behaviors. *Journal of Behavioral Education, 8,* 369–380. doi:10.1023/A:1022827415544.

Webb, M. S., de Ybarra, D., Baker, E. A., Reis, I. M., & Carey, M. P. (2010). Cognitive–behavioral therapy to promote smoking cessation among African American smokers: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 78,* 24–33. doi:10.1037/a0017669.

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral versus nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 317–320.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13,* 141–154. doi:10.1037/h0088978.

Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis, 39,* 369–373. doi:10.1901/jaba.2006.144-05.

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30,* 693–696. doi:10.1901/jaba.1997.30-693.

Accuracy and Validity of Methods for Identifying Learning Disabilities in a Response-to-Intervention Service Delivery Framework

Jeremy Miciak, Jack M. Fletcher and Karla K. Stuebing

Identification accuracy is a question of central importance in research and practice related to learning disabilities (LD) and other neurodevelopmental disorders. In research, any assessment of strengths and weaknesses in cognitive or neurological functions associated with LD or treatment outcomes depends on the validity of the criteria used to identify LD status. In practice, LD identification decisions have significant ethical and legal implications because LD status is related to intervention services, civil rights protections, and access to testing accommodations and modifications. A critical question is the reliability and validity of the criteria used to identify LD status. Despite the importance of reliable and valid identification decisions, finding agreement on the best methods for identifying LD status has proven elusive. This difficulty is not because of disagreement about the core attributes of LD, which most agree includes low, unexpected underachievement. LD status is exemplified by the person who seemingly has all the necessary attributes to learn to read, write, and do mathematics, but unexpectedly struggles to learn these skills (Fletcher et al. 2007).

J. Miciak (✉) · J. M. Fletcher · K. K. Stuebing
Department of Psychology and Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, TX, USA
e-mail: Jeremy.Miciak@times.uh.edu

Exclusionary Definitions

In the early history of identifying LD, it was assumed that if known causes of low achievement could be eliminated, such as an intellectual disability, sensory problems, economic disadvantage, or language minority status, those underachievers who remained could be identified with LD. Yet there are many problems associated with identification by exclusion, including the difficulty of defining the relations of exclusionary criteria to achievement, the circular nature of such an approach to identification, and the heterogeneous group of low achievers remaining when presumably known causes of low achievement are identified (Ross 1976; Rutter 1982; Satz and Fletcher 1980). For example, Taylor et al. (1979) identified poor readers who met and did not meet a variety of exclusionary criteria. The researchers were not able to distinguish poor readers identified as LD based on the absence of exclusionary factors from those identified as not LD on a variety of external cognitive, achievement, and brain status measures. This study epitomizes the failure of definitional approaches to older, neurological conceptions of LD based solely on exclusionary criteria, such as minimal brain dysfunction. However, ensuring that other exclusionary criteria are not the primary cause of low achievement remains an important feature of virtually every definition of LD.

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_25

Cognitive Discrepancy Definitions

Subsequent attempts to develop LD identification approaches focused on establishing inclusionary criteria for low achievement and to operationalize the "unexpectedness" component, which was essential for distinguishing LD from other forms of low achievement. The most prominent efforts involved aptitude–achievement methods that identified LD status when achievement was well below levels predicted by different kinds of intelligence quotient (IQ) tests, listening comprehension measures, or through a pattern of cognitive processing strengths and weaknesses (PSW). These approaches, which represent the most prominent methods for the identification of LD for the past 40 years, have also failed because such methods cannot reliably differentiate people with low achievement who demonstrate cognitive discrepancies from those who do not (Fletcher et al. 2007).

The validity problems of identification based on discrepancies of IQ and achievement are well known. Fundamentally, it is difficult to differentiate groups of low achievers with and without a discrepancy between IQ and achievement on variables not used to define the groups. This difficulty extends not only to external measures of achievement and cognitive function (Fletcher et al. 1994; Hoskyn and Swanson 2000; Siegel 1992; Stanovich and Siegal 1994; Stuebing et al. 2002) but also to treatment response (Stuebing et al. 2009; Vellutino et al. 2006) and brain function (Tanaka et al. 2011; Simos et al. 2013). It is important to note that these methods are not invalid because individuals with an IQ–achievement discrepancy are not LD. In fact, many individuals demonstrating an IQ–achievement are LD, because they meet definitional criteria, including low achievement. However, the underlying classification lacks validity because cognitive discrepancies are not valid markers of unexpected underachievement.

Even if cognitive discrepancy were a valid marker of unexpected underachievement, the identification criteria generated by these definitions do not permit reliable identification of LD status. This is because of underlying psychometric factors that will plague any approach to identification based on correlated attributes, such as achievement and aptitude. Some of these factors are specific to cognitive discrepancy methods, related to the use of difference scores on correlated attributes (Bereiter 1967). Difference scores involving a measure of aptitude (whether aptitude is measured as IQ, listening comprehension, or another cognitive process) and achievement possess the unreliability of both correlated measures. Thus, the reliability of any discrepancy approach is inherently lower than an approach based on a single test, despite their historical prominence and the frequency with which discrepancy approaches are proposed and accepted without strong empirical bases.

Unfortunately, the problem of unreliability in identification is not solved by abandoning difference scores. This problem is apparent for approaches based on low achievement, even when utilizing measures of similar constructs with the same cut point, because different decisions about who demonstrates low achievement will inevitably emerge as a result of measurement error (Francis et al. 2005). Moreover, these problems affect efforts to identify LD status based on recently proposed methods for LD identification: those based on evidence of inadequate response to quality instruction.

The fundamental message of this chapter is that low reliability for individual decisions about LD status is a universal problem with methods based on setting a cut point as a firm threshold on a univariate or bivariate distribution at a single point in time. Some approaches, such as those based on low achievement or instructional response, show validity at the level of the underlying classification. Adequate and low achievers, as well as adequate and inadequate responders to instruction, can be differentiated on a wide variety of attributes not used to identify low achievement or inadequate response, which is the cardinal characteristic of a valid classification. Nonetheless, a single measure of achievement or instructional response used to assess a person's status in relation to a firm threshold is inadequately reliable when identifying individuals as LD. Other approaches, such as those based on some form of cognitive discrepancy, lack validity at the group level because the underlying classification lacks validity.

In the remainder of this chapter, the conceptual framework essential for understanding identification accuracy is briefly reviewed, which is anchored in empirical classification research. Then, the chapter focuses specifically on methods that incorporate instructional response as the newest approach to LD identification. To illustrate the universal nature of identification problems for LD, the reliability and validity of identification approaches based on simple low achievement, IQ–achievement discrepancies, and PSW are discussed. The chapter concludes with a discussion of possible solutions to the difficulties with LD identification.

Classification, Definition, and Identification

The difficulties presented by cognitive discrepancy methods illustrate the conceptual framework essential for understanding issues related to identification accuracy. The criteria used to identify LD status represent an operationalization of a definition of LD. This definition stems from a conceptualization of LD and specifically how unexpected underachievement can be indexed. The conceptualization of LD thus specifies what attributes indicate LD and what attributes do not indicate LD. The definition is based on the attributes of LD, which are then operationalized as a set of identification criteria. Thus, classification, definition, and identification are not interchangeable, but represent different levels of assessment (Morris and Fletcher 1988).

When a study compares individuals identified as LD and not LD, the study usually represents an explicit comparison on some other attribute, such as brain function, that is believed to be different in those identified with and without LD. If a study showed differences in brain function, it is not only an explicit validation of a hypothesis about how brain function is associated with LD status but also an implicit validation of the identification criteria, the definition, and the classification that generated the definition. Thus, a classification model may posit that an essential attribute of unexpected underachievement is a cognitive discrepancy. Stemming from this *hypothesis* is a definition that stipulates that LD is indicated by a severe discrepancy between a measure of aptitude and achievement. For identification, the definition is operationalized by specific criteria for determining the size of the discrepancy between aptitude and achievement necessary and the specific IQ and achievement tests to be utilized. Simply demonstrating that people possess the specified IQ–achievement discrepancies, or that if they do so they struggle in school, is not an evaluation of the underlying cognitive discrepancy classification. It is only when the classification hypothesis is explicitly formulated and tested against counterfactual specifying attributes that do not represent LD that tests of validity are possible. Similarly, finding that people have strengths and weaknesses in cognitive processing or that some people do not respond to instruction does not test the classification hypothesis and does not demonstrate that the underlying classification is valid. Again, it is only when a counterfactual is formulated and tested by comparing members who vary in the definitional attributes (e.g., poor readers with and without cognitive discrepancies; adequate and inadequate responders to instruction) *on variables not used to identify status* that the validity question is addressed. Obviously, groups identified with and without PSW, or adequate and inadequate responders, will differ on the definitional attributes unless they are completely unreliable; external validity accrues only if the groups differ on variables not used to define the groups.

LD Is a Latent Construct

Moving towards the issue of identification accuracy, it is important to recognize that much of the ensuing discussion is really about definitions and identification criteria. However, definitions and criteria are derived from a classification framework that hypothesizes how LD is indicated and not indicated. LD itself is an unobservable (latent) construct that can only be known by how it is measured, which is why recognizing the implicit nature of the underlying classification is important.

Recognizing that identification of LD is unreliable, or that a particular classification hypothesis lacks validity, does not mean that the construct of LD is not real. Such problems indicate that the measurement process lacks validity. Indeed, it is easy to identify people who possess the attributes of LD and to show that people with low achievement can be differentiated from people with adequate achievement on measures not utilized in the identification process, thus meeting the criteria for classification validity outlined above (Fletcher et al. 2007). *LD is real*; our efforts to measure this latent construct are often not adequate. Evaluations of the identification accuracy of different approaches to identify LD are complicated because there is no “gold standard” against which observed identification decisions can be compared. This is not unique to LD. Rather, it characterizes many psychological and medical conditions. LD represents a latent construct, unobservable outside of attempts to measure definitional attributes.

This seemingly esoteric distinction between measured and latent levels has important implications for the accuracy of all approaches to identify LD. First, and most importantly, all attempts at the identification of LD will be subject to error because the measures utilized to assess attributes of LD are imperfect. All measures, whether they assess cognitive processing, executive functioning, or more familiar constructs such as reading or mathematics, are imperfectly reliable and correlate imperfectly with the latent construct of interest. This measurement error impacts the reliability and validity of resulting classifications (Francis et al. 2005). Additionally, because of the latent nature of LD, researchers and practitioners often conflate observed attributes with the unobservable disorder. This creates confusion in classification research and often results in circular reasoning, in which a specific pattern of assessment results is treated as the disorder itself. LD represents an unobservable, latent classification. Definitions of LD attempt to establish the inclusionary and exclusionary criteria that would demarcate the disorder and lead to an identification process that can operationally define these criteria through observed performance. Yet, this operational definition should be recognized as an imperfect attempt to measure a latent construct.

The Attributes of LDs Are Dimensional, Not Categorical

In practice, LD status is treated as a dichotomous, categorical determination. A student is classified as either LD or not. That classification is based on the presence or absence of definitional attributes and measured through the chosen identification process. However, it is important to note that this categorical determination reflects practical and legal exigencies rather than intrinsic characteristics of LD. Indeed, the attributes of LD appear to be normally distributed, with no discontinuities that would mark naturally occurring groups (Fletcher et al. 2007). While it is common to interpret LD within a medical disease model, in which individual status can be positive or negative (similar to diagnoses for infectious disease or other categorical classifications), there is little evidence to support this understanding of LD. A better medical analogy is found in dimensional conditions such as obesity or hypertension (Ellis 1984). Consider hypertension, which is indicated by consistently elevated systolic and diastolic blood pressure. While clinical guidelines specify a normal range and thresholds for elevated risk, there is no naturally occurring line which would differentiate individuals with hypertension from those without.

The attributes of LD present similarly. Across studies investigating attributes of LD, including academic achievement, response to instruction, or even cognitive functions, there is no demarcating line that would separate students with LD and students without LD into qualitatively different groups. Instead, the distribution of these attributes appears continuously distributed when students with intellectual disabilities are excluded (particularly students with acquired conditions related brain injury). This is most clearly illustrated in population studies of academic achievement, in which student performance in reading and mathematics are normally distributed (Jorm et al. 1986; Rodgers 1983; Shaywitz et al. 1992; Silva et al. 1985). In recent years, large-scale studies investigating response to reading intervention have further supported the dimensional nature of attributes of LD, as individual instructional response is continuously distributed, without naturally occurring demarcations of qualitatively different groups (Fletcher et al. 2011; Miciak et al. 2014a; Vellutino et al. 2006). Further, studies applying methods from behavioral genetics have failed to identify qualitatively different genetic constellations associated with the heritability of reading and mathematical disorders (Fisher and DeFries 2002; Grigorenko 2001, 2005; Plomin and Kovas 2005).

The dimensional nature of the attributes of LD has important implications for the identification of individual students with LD. First, it is conceptually inconsistent to apply a categorical structure to continuous data. Indeed, methodologists have long criticized the practice of subdividing a continuous distribution to create groups (Cohen 1983). Any attempt to subdivide a continuous distribution is inherently arbitrary because there are no naturally occurring, qualitatively different groups. Individuals close to the cut point perform similarly on both the criterion construct and other correlated constructs. Additionally, applying a group structure to continuous data reduces “within group” variability and reduces the range of measurement. This reduces the expected correlations with other variables. The power to detect relations among dependent and independent variables is reduced, and inaccurate results may emerge due to the failure to control fully for the correlation between the two dimensions categorized (MacCallum et al. 2002).

The underlying arbitrariness of the imposed group structure is highlighted by the observed unreliability of identification decisions when cut points are applied. Scores for repeated testing will fluctuate around the cut point, creating unreliability in identification decisions (Francis et al. 2005; Macmann et al. 1989). This fluctuation is partly due to the measurement error of any psychometric test, including tests of cognition, executive functions, or academic achievement, and is inherent to any attempt to dichotomize performance utilizing continuous, imperfect measures.

Issues related to unreliability due to measurement error and the imposition of an arbitrary group structure to continuous data are well documented and understood (Fletcher et al. 2012). However, these issues have not been adequately addressed in LD research and public policy. Indeed, discussions about LD classification, definition, and identification tend to focus on the latent attributes that constitute the disorder and rarely address these underlying psychometric issues (see for example, Hale et al. 2010). To the extent that unreliability due to underlying psychometric limitations is addressed, it is often to critique specific approaches to LD identification and ordinarily ignores the universality of these issues. An alternative is to move towards an approach that incorporates multiple inclusionary criteria, utilizes confidence intervals for decision-making, and shifts the identification process towards an evaluation of risk and/or probability of academic difficulty, rather than an actuarial process aimed at identifying the right children. This shift may begin to address the methodological issues that are highlighted in the remainder of this chapter.

Response to Instruction and LD Identification

The idea of identifying LD based on indicators of instructional response achieved prominence in the most recent reauthorization of IDEA, although this idea was firmly embedded in research for many years preceding this reauthorization (Fuchs and Fuchs 1998). The explicit inclusion of methods of LD identification that incorporate instructional response is often discussed as a fundamental shift in procedures, which is incompatible with previous methods of LD identification. In fact, many of the provisions upon which LD eligibility in IDEA 2004 is based are universal and not specific to identification within response-to-intervention (RTI) models. Across allowed methods, LD identification requires: (a) a comprehensive assessment, (b) evidence of (limited) academic progress in the general education curriculum, (c) consideration of the appropriateness of the general education curriculum, (d) consideration of exclusionary criteria, including other disorders and/or factors that may explain academic underachievement, and (e) the consent of a parent or legal guardian. None of these requirements change when LD identification is situated within an instructional response model. Indeed, to refer to LD identification based upon RTI is an oversimplification because, in policy, LD identification has never been based on a single criterion.

Definitions of LD that incorporate instructional response as an inclusionary criterion should be conceptualized as a different classification hypothesis concerning the intrinsic nature of LD. In these approaches, the underlying, implicit conceptual model involves an inability to respond to quality instruction that works with most people, which indicates that low achievement is unexpected. Thus, classifications that incorporate instructional response as the marker of unexpected underachievement lead to different definitions and identification criteria from methods that conceptualize unexpectedness as a cognitive discrepancy. In particular, because instructional response is an inclusionary criterion within an instructional model of LD and exclusionary criterion in a cognitive discrepancy model of LD, the former incorporates different measurements. Instructional models identify LD as low achievement and inadequate response to instruction. They do not specify a cognitive discrepancy as necessary or sufficient for identification. Thus, cognitive processes would not be assessed in an instructional model (except to address exclusionary criteria, such as an intellectual disability); they would measure achievement and instructional response because they are inclusionary criteria.

To illustrate a definition of LD based on an instructional model, consider the results of the 2001 LD Summit convened by the United States Office of Special Education. At the end of this summit, the Office of Special Education convened a group of prominent LD researchers and stakeholders for a consensus meeting on the definition and identification of LD (Bradley et al. 2002). From that meeting, there emerged consensus on a three-criterion definition of LD, including two inclusionary criteria and one exclusionary criterion. In order to be identified as LD, the interdisciplinary team must document three components: (a) low academic achievement; (b) inadequate response to research-based, generally effective interventions; and (c) absence of exclusionary factors, such as sensory deficits, intellectual disabilities, English language learner status, or emotional disturbance that *primarily* cause low achievement. Importantly, this definition was not developed specific to an approach based on RTI service delivery models. Rather, these three criteria were deemed necessary to identify LD regardless of the adopted approach. However, an LD identification method based, in part, upon documenting inadequate RTI may be well situated to address each of these three criteria. Moreover, this definition and the underlying classification framework can be aligned with the policy requirements of IDEA 2004.

The Reliability and Validity of Methods Based on RTI Service Delivery Methods

There is considerable confusion regarding the relation of RTI service delivery models and methods of LD identification based, in part, on an assessment of instructional response. This confusion exists because many critics reduce RTI service delivery models to a process that exists for LD identification only. However, the primary focus of RTI service delivery models has never been LD identification. Instead, the RTI service delivery model is designed to provide timely, evidence-based academic interventions and enhance learning outcomes for all students, including students with potential LDs (Fletcher and Vaughn 2009). In the sections that follow, issues related to the assessment of intervention response, the component of the identification process that is derived from an RTI service delivery model, are reviewed. However, it is worth noting once more that LD identification has never been based on a single data point, whether it is an index of intervention response, a score documenting low achievement, or testing results that confirm a specific pattern of cognitive PSW.

Several considerations are important for understanding the studies that will be reviewed. The chapter focuses on the accuracy of methods for identifying individual LD status. The simplest evaluation of this type of accuracy is to apply proposed criteria and compare resulting identifications to known "true" LD status. The percentage agreement provides a powerful, easily interpretable statistic through which potential approaches to identification can be evaluated. Such evaluations are common in medical research, in which true status is often readily ascertained through a blood test or medical imaging. However, this is not possible in observed data because there is no "gold standard" for LD status and any decision about "true LD" is somewhat arbitrary because of the dimensional nature of the attributes. An alternative is to compare agreement across methods. Reliable identification methods would be expected to show agreement across similar methods. Most studies calculating agreement across methods report a chance-corrected index of agreement, such as kappa (κ, Cohen 1960). Cicchetti and Sparrow (1981) proposed an interpretive framework for κ, in which $\kappa > 0.75$ is considered excellent, 0.60–0.74 good, 0.40–0.59 fair, and < 0.40 poor. By consensus, $\kappa < 0.40$ is considered undesirable.

A second important consideration involves coverage, which is the number of people identified with LD relative to an arbitrary standard (e.g., a pool of inadequate responders or all the children identified with LD in a school). A reliable identification method should identify 70–80 % of this type of pool; in a statistical simulation, this could be established by the prevalence, or base rate, of LD, although this criterion is also somewhat arbitrary because of the multiple factors that may affect coverage.

It is also important to consider the nature of decision errors because the relative effects of a false-positive error (identifying a person as LD when their "true" status is not LD) or a false-negative error (missing a "true" LD) may not be equal. In this case, the percent positive value and percent negative value may be considered to determine the portion of individuals correctly identified as having the condition and the portion correctly identified as not having the condition. This allows researchers and policymakers to consider the costs and benefits associated with the proposed identification approach. There are no firm guidelines for acceptable false-positive and false-negative errors. The criteria are ultimately pragmatic, based on achieving an optimal balance between high sensitivity (the ability to correctly identify individuals with the condition) and high specificity (the ability to correctly identify individuals without the condition). For example, in an early screening situation, a failure to provide services when needed (false negative error) is more deleterious than a false-positive error, the latter affecting resource allocation, but not depriving a person of needed intervention.

Reliability Issues Related to Determining Intervention Response

Intervention response, like all attributes of LD, is an unobservable, latent construct. Recent large-scale studies suggest that intervention response lies on a continuum, with no naturally occurring demarcations which would separate adequate from inadequate responders (Fletcher et al. 2011; Miciak et al. 2014a; Vellutino et al. 2006). Thus, students whose measured performance lies near the cut point(s) (however it is defined) are likely to be very similar, and group membership will fluctuate across different measures or measurement occasions. Further, because intervention response represents a complex, latent construct, how it is operationalized will vary across different states and districts, as well as its operationalization in empirical research. Fuchs and Deshler (2007) identified five methods to operationalize inadequate response, including: (a) a median split, defined as a measured slope below the median; (b) normalization, defined as an observed score below a threshold (typically the 25th percentile) on a norm-referenced assessment; (c) final benchmark, defined as performance below a benchmark criterion-referenced assessment; (d) dual discrepancy, defined as slope and final level at least one SD below that of peers; and (e) slope discrepancy, defined as slope below a normative cut point. In broad terms, these methods rely on either student growth over time, post-intervention performance, or both (dual discrepancy). However, as evidenced in empirical studies, none of these methods to determine individual intervention response overcomes the psychometric limitations associated with imposing an arbitrary, dichotomous structure on continuous data derived from imperfect measures. Although the focus of critics of approaches emanating from RTI frameworks has been on curriculum-based methods for assessing progress, it is important to recognize that many studies use traditional norm-referenced tests of achievement as a final status method.

Empirical Studies Two studies have reported agreement for inadequate responder status achieved by methods based on growth, final status, and dual discrepancy (Barth et al. 2008; Brown-Waesche et al. 2011). Barth et al. investigated differences in identification status resulting from differences in cut points, tests, and methods utilized to identify inadequate responders. All 399 first-grade participants completed a tier 2 intervention or participated in enhanced tier 1 instruction (Mathes et al. 2005). The measures for determining intervention response included two assessments incorporating growth, thus permitting calculation of slope-only and dual-discrepancy methods. Additionally, four final status measures assessed decoding, fluency, and comprehension in reading. Cut points for discrepant status included 0.5, 1, and 1.5 standard deviations (SD) below the normative mean. Over 808 total comparisons, only slightly over 15 % achieved a minimum level of agreement ($\kappa < 0.40$). Further, the comparisons that yielded acceptable agreement were driven largely by agreement on adequate responders, rather than high agreement on inadequate responders (Barth et al. 2008).

Brown-Waesche et al. (2011) compared agreement among four definitions of reading disability, including a low achievement definition and three definitions based on intervention response: (a) final status, (b) slope-only, and (c) dual discrepancy. The sample consisted of 288,114 students in first through third grades. Each of the definitions of reading disability was evaluated at six cut points, ranging from the 3rd to the 25th percentile. Results yielded poor agreement between the low achievement and intervention response definitions of reading disabilities. Comparisons of methods based on the intervention response also yielded poor to fair agreement both across and within methods. Similar to Barth et al. (2008), agreement was higher for identifying adequate responders with lower agreement for determinations of inadequate response.

Burns et al. (2010) evaluated agreement and internal consistency for decisions about the adequacy of response using a dual-discrepancy method and a method based on progress toward a final benchmark (aim-line approach). The two methods were applied to determine if students were making inadequate progress, adequate progress, or exceeding expectations. Comparisons of the categorical decisions of the two methods (aim line and dual discrepancy) yielded poor agreement ($\kappa = 0.29$). Further, when data were compared in split sets to evaluate the internal consistency of categorical decisions, both methods yielded coefficients in the poor to fair range ($\kappa = 0.59$ for dual discrepancy; $\kappa = 0.29$ for aim line), indicating that low agreement is not merely a result of applying different methods for determining adequate response.

It is important to note the observed low agreement does not simply reflect limited agreement across different methods for determining intervention response (i.e., final status vs. dual discrepancy or slope-only). The studies above documented poor to fair agreement even when the same method to determine intervention response was utilized (Barth et al. 2008; Brown-Waesche et al. 2011) and poor internal consistency for identification decisions (Burns et al. 2010). Poor agreement is also documented in studies that evaluate different ways to operationalize a single method, such as a dual-discrepancy method. Burns and Senesac (2005) evaluated four definitions of dual discrepancy, in which inadequate response was defined as an end-of-year oral reading fluency score indicating risk and growth below one of four normative benchmarks (the 25th percentile, 33rd percentile, 50th percentile, and >1 SD below mean growth). Identification status fluctuated considerably, illustrating the fundamental importance of cut points in defining group membership. Across cut points for growth, the portion of inadequate responders ranged from 12.3 % of the sample identified with a cut point of 1 SD below mean growth to 41.8 % of the sample for the 50th percentile cut point.

Simulation Studies To illustrate the underlying psychometric issues that may explain the low agreement observed in empirical studies, Fletcher et al. (2014) simulated data based on two highly reliable final status measures with manipulations of reliability, the intercorrelations of the two measures, constructs measured, cut points, normative samples, and sample size. For each condition except the small sample size condition, 100,000 simulated observations were generated that recreated the condition parameters (e.g., the specified reliabilities, intercorrelations of the two measures, cut points, normative samples, and sample size for the specific condition). Simulating data in this way allows for systematic manipulation of variables of interest, permitting an evaluation of the effect of specific variables on subsequent identification decisions.

If the final status tests were correlated at 1.0 (which would mean both had perfect reliability), the agreement would be 1.0. For two normally distributed, perfectly measured variables with a correlation of 0.94 (disattentuated for the unreliability of the two tests) and a cut point <25th percentile, agreement was excellent ($\kappa = 0.76$). If the correlation between the measures was perfect, but the reliability was reduced to the reported reliabilities of two highly reliable norm referenced measures, 0.98 for the Woodcock Johnson III Tests of Achievement (Woodcock et al. 2001) and 0.90 for the Test of Word Reading Efficiency, (Torgesen et al. 1999), κ was again 0.76. When the construct-level correlation was 0.94 and published reliabilities were used, κ ranged from 0.62 to 0.68. However, when the simulation was adjusted to reflect normative sample differences and small sample size, κ was only poor to fair ($M = 0.39$; range = 0.27–0.51). Notably, this limited agreement occurs in a simulation of two highly reliable, norm-referenced measures of reading achievement. Although critics of RTI methods often criticize identification methods based on RTI because of methodological difficulties related to identifying inadequate response (Reynolds and Shaywitz 2009; Hale et al. 2010), these methodological difficulties result from the use of two reliable norm-referenced achievement tests. Imagine the results measures of IQ and achievement were simulated, or two cognitive process measures.

Validity of Intervention Response as an Attribute of LD

The reliability issues discussed above concern LD identification based on intervention response at the individual level. Because of the psychometric limitations that have been highlighted, no identification process will demonstrate perfect reliability. However, this imperfect reliability does not indicate that the overall classification is invalid. For classifications based on intervention response, inadequate and adequate responders should be compared on measures not utilized to identify response status. These measures could assess highly related attributes, such as reading or writing, or could assess more distal attributes such as behavior, attention, or even subsequent intervention response. At this level, empirical studies suggest that classifications based on differential intervention response consistently separate groups on a number of theoretically related attributes, including academic level (Al Otaiba and Fuchs 2006; Nelson et al. 2003; Vellutino et al. 2006), cognitive characteristics (Fletcher et al. 2011; Miciak et al. 2014a), behavior (Al Otaiba and Fuchs 2006; Nelson et al. 2003), and even brain activation patterns (Molfese et al. 2013; Rezaie et al. 2011).

Such data provide evidence for the validity of intervention response as a classification attribute because they suggest that subgroups can be separated across attributes not utilized for group formation. However, differences in level on related attributes should not be interpreted as evidence for a categorical disorder. Instead, findings are generally consistent with a continuum of severity hypothesis (Vellutino et al. 2006), in which achievement and achievement-related abilities lie on a continuum reflecting the severity of the achievement difficulty. The fundamental question is whether group separation between adequate and inadequate responders reflects differences in the *level* of performance or differences in the *pattern* of performance. A difference in pattern would suggest that the subgroups are qualitatively different, consistent with a categorical disorder. Fletcher et al. (2011) investigated this issue in a study of the academic and cognitive attributes of adequate and inadequate responders to a tier 2 reading intervention. Groups of inadequate responders were identified using fluency and decoding final status measures, yielding groups with deficits in decoding and fluency, fluency-only, and adequate responders as well as typically achieving children. Across measures evaluating phonological awareness, rapid naming, expressive and receptive language, working memory, nonverbal problem-solving, and vocabulary/verbal knowledge, a clear hierarchy emerged, with the typically achieving and adequate responder groups outperforming both inadequate responder groups. Multivariate statistical tests were generally insignificant when comparing the two inadequate responder groups, but both groups were statistically different from the adequate responder and typically achieving group.

Similarly, Miciak et al. (2014a) investigated the cognitive attributes of adolescent adequate and inadequate responders identified by final status indicators in decoding, fluency, and comprehension. Application of adequate response criteria (SS > 90) yielded four subgroups: (a) adequate responders (SS > 90 on all measures), (b) poor comprehenders (SS < 90 on Woodcock Johnson-III Passage Comprehension only), (c) impaired fluency (SS < 90 on TOWRE composite), and (d) globally impaired readers who scored below the cut point on all criterion measures. Results indicated a clear hierarchy on both cognitive and academic variables, with the adequate responder group scoring highest on academic and cognitive variables, the groups with specific reading deficits (comprehension or fluency) scoring similarly, and the globally impaired group scoring lowest on both cognitive and academic measures. In both of the above studies, the differences in level observed on cognitive and academic variables between adequate and inadequate responders parallel differences in readings skill, consistent with a continuum of severity hypothesis.

Alternative Approaches for LD Identification

Thus far, classification, definition, and identification issues have been discussed primarily within the context of LD identification methods situated within RTI service delivery models and utilizing intervention response as an inclusionary criterion for LD identification. A central theme highlighted the universality of the methodological issues that emerge within this classification framework. In the sections that follow, alternative approaches for LD identification and issues related to the reliability and validity of each approach are discussed.

Approaches Based on Low Achievement

Reliability One approach to LD identification focuses on low achievement as the primary inclusionary criterion for LD identification (Siegel 1992). This direct approach avoids many of the complications associated with approaches based on assessing discrepancies because single measures are involved. However, low-achievement methods do not resolve the psychometric issues involved in identifying individual people with LD and highlight the difficulties associated with applying rigid cut points (Francis et al. 2005). Instead, the identification status of individual students will demonstrate instability, particularly for students close to the cut point because of imperfect reliability and imperfect measurement of the latent construct.

Francis et al. (2005) illustrated this instability in identification decisions utilizing simulated and longitudinal data. Data from grades 3 and 5 were selected for participants in the Connecticut Longitudinal Study (CLS). Assessment data in grade 3 were utilized to identify students who were typically achieving, low achieving, discrepant (i.e., demonstrated IQ–achievement discrepancy), or discrepant and low achieving. Grade 5 results were then used to evaluate the stability of group membership for individual students. Notably, only 63 % of students in the low-achieving (but not IQ discrepant) group in grade 3 remained in the low-achieving group at grade 5. Further, only 68 % of the low-achieving group maintained LD status (defined as low achievement, IQ discrepant, or both). This finding was replicated using simulated data, which found only 51 % of the low-achieving (but not IQ discrepant) group in grade 3 remained in the low achieving at grade 5 and only 64 % of the low-achieving group maintained LD status. Such findings highlight the instability of applying rigid cut points in one point of time.

Another problem that emerges when considering low-achievement methods is determining what level of achievement would serve as the threshold for LD identification. In studies of reading, it is common to utilize a cutoff of the 25th percentile. This cut point is not universal across empirical research. Some mathematical studies set a more restrictive cut point of performance below the 10th percentile (Geary et al. 2007) to qualify as LD in mathematics. Such decisions are ultimately arbitrary, but have important implications for the prevalence of LDs. If, for example, LD in reading is defined as achievement below the 25th percentile with considerations of exclusionary clauses, 2 % may be excluded because of intellectual disabilities and perhaps another 3 % for other exclusionary clauses. The resulting 20 % prevalence is likely much higher than most would consider optimal. However, the assertion that this prevalence is unmanageably high is based on historic, political, and economic considerations rather than any scientific evidence, per se. Regardless, this issue affects every proposed method for identifying LD.

Validity It is well known that groups defined with low achievement in reading and mathematics can be differentiated on measures of cognitive and brain function from typical achievers, from people with intellectual disabilities, and from children with other disorders where achievement is not directly impacted (e.g., attention-deficit hyperactivity disorder (ADHD) with no comorbid LD). In addition, there are clear differences in the heritability of reading and mathematical skills, although the area also shared genes across different forms of LD (Plomin and Kovas 2005). Most importantly, validity of low-achievement classification accrues from treatment studies. For example, studies of reading interventions find that the reading performance of poor readers improves when they receive reading instruction but does not improve when they receive additional mathematical instruction (Morris et al. 2012). Even within specific academic domains, an assessment process that identifies specific skill profiles demonstrates good treatment validity. Connor et al. (2009) evaluated the decoding and comprehension skills of 461 first graders to inform an intervention. Intervention teachers were assisted in adjusting the amount and focus of students' word reading and reading comprehension instruction by a computer algorithm. Results indicated that close adherence to recommended amounts of individualized instruction improved student outcomes, providing strong evidence for aptitude by treatment interactions *when the treatment is tailored to academic needs.*

The most significant issue for the validity of low-achievement approaches is that it strays from the historic notion of "unexpected underachievement" and fails to meet the three-criterion consensus definition of LD (Bradley et al. 2002). It is easy to imagine a low-achievement identification process that includes consideration of exclusionary clauses, similar to other proposed approaches. However, an approach with low achievement as the only inclusionary criterion would not provide evidence of inadequate response to evidence-based, generally effective interventions. It is therefore out of step with mainstream definitions and conceptions of LD. Yet this critique in no way impacts the reliability and validity of the proposed classification framework. That framework represents a hypothesis which is subject to empirical validation. In the end, it may be that low-achievement approaches represent a sufficiently reliable, highly valid classification that is not desirable for widespread implementation for historic and political reasons.

Approaches Based on Aptitude–Achievement Discrepancies

Reliability Because of the dimensional nature of LDs and psychometric limitations associated with dichotomizing continuous data based on specific cut points, IQ–achievement methods for LD identification demonstrate poor reliability for the identification of individual students (Francis et al. 2005). Indeed, to reiterate, issues of reliability are exacerbated because a cut point on a single discrepancy score incorporates measurement error from two measures (Bereiter 1967). Additionally, the reliability of discrepancy methods is affected by the correlation between the two measures. Thus, discrepancy approaches should take into account the correlation of the two measures (Macmann and Barnett 1985). If a simple difference score is utilized, comparisons are subject to regression to the mean, such that individuals with a high aptitude score will be overidentified as LD, while individuals with a lower aptitude will be underidentified as not LD (Fletcher et al. 2012).

The psychometric issues underlying discrepancy approaches to LD identification are well documented. MacMann and Barnett (1985) identified three factors which impact the reliability of identification within a discrepancy approach: (a) the reliability of the difference, because difference scores are less reliable than single test scores, (b) test selection, and (c) the cut point utilized to determine a severe discrepancy. Using simulated data, Macmann and Barnett compared identification rates for discrepancy methods that varied in the reliability of the measures, the intercorrelation of the measures, and the cut point utilized. Across simulations, the agreement was unacceptably low, achieving between 50 and 60 % agreement only under the most optimal circumstances. Macmann et al. (1989) extended this work with a sample of 373 students and simulated data to determine if the method by which the IQ–achievement discrepancy is calculated influenced identification decisions and the impact of different academic measures. Across different calculation methods, the agreement was fair to excellent (κ ranged from 0.57 to 0.86). However, when different achievement measures within the same academic domain were utilized, agreement on identification status dropped substantially, with κ ranging from 0.19 to 0.47 (poor to fair agreement).

Francis et al. (2005) also evaluated the stability of identification decisions based on an IQ–achievement discrepancy over time. The study utilized observed IQ and achievement data in grades 3 and 5, as well as data simulated assuming high stability (0.90) and high reliability (0.80). The data recreated a bivariate distribution of IQ and achievement data with an intercorrelation of 0.60. Evaluating the identification decisions that result in both simulated data and actual data from the CLS, Francis et al. reported that over 30 % the children identified as LD or not LD in grade 3 change status by virtue of a repeated assessment. Such findings highlight the inherent volatility of actuarial decisions based on a single data point, in this case a discrepancy score.

Validity At present, few researchers concerned with LD would argue for the validity of classifications based on an IQ–achievement discrepancy. Two recent meta-analyses have compared subgroups of struggling readers with and without an IQ–achievement discrepancy to investigate whether the emergent groups differ in their behavioral, cognitive, and academic performance (Hoskyn and Swanson 2000; Stuebing et al. 2002). Both meta-analyses found little support for qualitative differences between groups of struggling learners that met or did not meet the IQ–achievement discrepancy. Subsequent studies have further demonstrated that IQ and IQ–achievement discrepancies are minimally predictive of treatment response (Stuebing et al. 2009; Vellutino et al. 2000) and struggling readers with and without an IQ–achievement discrepancy do not differ in their rate of reading growth or their level of reading ability at any age (Francis et al. 1996). Finally, functional brain imaging studies have failed to identify differences in the brain activation patterns of children with and without an IQ–achievement discrepancy (Simos et al. 2013; Tanaka et al. 2011).

Approaches Based on PSW

Cognitive approaches to LD identification based on an IQ–achievement discrepancy are largely discredited. However, in recent years, there have been increasing calls for methods of identification based on an assessment of cognitive PSW. These approaches hypothesize that LD is marked by a specific pattern of cognitive strengths and weaknesses combined with specific academic weaknesses (Flanagan et al. 2007; Hale and Fiorello 2004; Naglieri 1999). Proponents of these approaches make strong evidentiary claims. Indeed, a recent white paper by the Learning Disabilities Association of America, which claimed to represent an expert consensus on issues related to RTI, comprehensive evaluation, and LD identification, concluded that an approach based on identifying a pattern of cognitive PSW "makes the most empirical and clinical sense" (Hale et al. 2010, p. 228). However, this conclusion is controversial and may not represent a consensus view among experts in the field of LD research, law, and practice (The Consortium for Evidence-Based Early Intervention Practices (CEBEIP) 2010). Further, despite claims to the contrary, empirical evidence in support of the reliability and validity of PSW approaches has been slow to emerge.

Three approaches have been proposed to operationally define LD identification via a PSW approach: (a) the concordance/discordance method (C/DM; Hale and Fiorello 2004), (b) the cross-battery assessment method (XBA; Flanagan et al. 2007), and (c) the discrepancy/consistency method (D/CM; Naglieri 1999). These methods differ in the way that inclusionary criteria such as low achievement and a profile of PSW are defined, as well as in how exclusionary factors are considered. For example, the C/DM represents an ipsative approach, in which observed cognitive scores are utilized to identify an intraindividual pattern of strengths and weaknesses. In contrast, the XBA is primarily a normative approach, in which the presence or absence of normative deficits is utilized to establish a profile of strengths and weaknesses, while the D/CM utilizes both ipsative and normative comparisons. The methods also differ in their theoretical orientation. For example, the XBA approach utilizes the Cattell–Horn–Carroll (CHC) theory of intelligence, while the D/CM prefers the planning, attention, simultaneous, and successive (PASS) factors of intelligence measured by the Cognitive Assessment System (Naglieri and Das 1997) and the C/DM emphasizing flexibility across measures and theoretical orientation.

Reliability None of these approaches overcomes the psychometric issues related to discrepancy scores, including the reliabilities and intercorrelations between the tests and fluctuations around cut points. The application of PSW methods will result in significant psychometric difficulties because all methods rely on multiple imperfect, correlated tests to establish a complex pattern of cognitive processing and academic achievement.

Stuebing et al. (2012) used simulation techniques to investigate the three proposed PSW approaches for LD identification. The simulation generated latent data based on multiple reliabilities, correlations between constructs, and cut points in accordance with the recommendations of each approach. Observed data were then generated using the reported reliabilities of tests. The classification agreement between simulated latent and observed classifications was evaluated. Results indicated that under all three methods, only a small number of students (1–2 %) met PSW LD identification criteria, raising questions about the efficiency of the model. All three approaches demonstrated high specificity and high negative predictive power, indicating that they are largely consistent in classifying students without LD. This is not surprising, given the low base rate of the disorder under these models. However, all three approaches demonstrated moderate to low sensitivity and very low positive predictive power, because false-positive rates for LD identification were high (Stuebing et al. 2012).

Miciak et al. (2014b) utilized observed data for a sample of 139 adolescents demonstrating inadequate RTI to empirically classify participants as meeting or not meeting PSW LD identification criteria according to the C/DM and XBA methods. An investigation of the resulting identification decisions permitted an analysis of the identification rates, agreement, and external validity of the two identification approaches. Both approaches identified a low percentage of participants as LD (range 24–66 %) in a sample limited to students demonstrating intractable reading deficits. Further, results indicated that the two approaches demonstrated poor agreement for LD identification decisions (κ range = −0.04–0.31), suggesting that the two approaches may not be interchangeable.

Validity Miciak et al. (2014b) also evaluated the external validity of the identification approaches by comparing the performance of participants that met criteria and participants that did not meet criteria on academic measures which were not utilized to identify groups. Comparisons failed to find meaningful differences in academic performance between emergent subgroups for either method. Resulting groups demonstrated similar patterns of academic performance which fluctuated according to the cut point for low academic achievement applied, rather than the identification of specific cognitive patterns. Such results raise questions about the validity of the underlying classification framework.

Another significant problem associated with PSW approaches to LD identification is the assumption that such approaches will improve treatment outcomes for children with LD (e.g., Hale et al. 2010). Proponents of PSW approaches contend that specific cognitive strengths and weaknesses may indicate different intervention approaches, which presumably provide training or support in areas of specific cognitive weaknesses. However, there is little evidence that instruction addressing specific cognitive processes is related to specific intervention outcomes (Fletcher et al. 2012; Reschly and Tilly 1999). There has been no shortage of studies searching for aptitude by treatment interactions in psychological and educational research. Despite this considerable attention, there remains little evidence to support the existence of aptitude by treatment interactions for cognitive patterns or learning styles (Kearns and Fuchs 2013; Pashler et al. 2009).

Some Potential Solutions

Empirical evidence suggests that the attributes of LD represent continuous distributions, with no naturally occurring discontinuities that would indicate qualitatively distinct groups. Yet, current policy and practice requires categorical decisions that arbitrarily dichotomize this dimensional attribute. As a result, all approaches to LD identification exhibit inherent psychometric limitations, because measures utilized to operationalize the attributes of LD are often intercorrelated and always demonstrate some level of unreliability. For individual students, this results in identification decisions with undesirable levels of unreliability. To a certain extent, this is inescapable unless conceptions of LD shift away from approaches that mandate high-stakes, categorical decisions which influence educational placement for years. With the reauthorization of IDEA in 2004, many states that shifted to methods incorporating instructional response still set cut points for low achievement and inadequate response that were sometimes different for the two attributes and had no basis except in arbitrary efforts to constrain resources. As the research reviewed in this chapter demonstrates, these methods will be inherently unreliable for individual decisions.

There are potential solutions to these problems. The first is to move away from rigid cut points and begin to employ confidence intervals that incorporate the measurement error of the tests. Confidence intervals emphasize the degree of uncertainty about a given test score rather than a point estimate by expressing performance as a range of plausible values. The degree of confidence desired might vary across applications but traditionally a 68 % confidence interval (CI) or 95 % CI is used. The first is created by setting the upper limit at 1 standard error of the mean (SEM) above the observed score and setting the lower limit at 1 SEM below the observed score. For a 95 % CI, the limits are obtained by finding the scores that are 1.96 SEM above and below the observed score.

A further modification to the typical use of confidence intervals is recommended to correct for the regression effects that are particularly likely in observations far from the mean (as is the case in students who are selected on the basis of low performance). In this modification, the confidence interval is centered around the estimated true score (Campbell and Kenny 1999) rather than the observed score. This approach is implemented in the Wechsler Intelligence Scale for Children, Fourth Edition (Wechsler 2003). The 95 % CI is formed by finding the scores that are 1.96 above and below the estimated true score rather than the observed score. The interpretation is that if this procedure were to be used to estimate a student's true score using a CI of the given width, the true score would be captured 95 out of 100 times.

The second potential improvement is to recognize that single assessments for LD status are not capable of reliability. One alternative would be to use multiple criteria and to set the cut points relatively high (or use confidence intervals) to avoid false-negative errors. Here, it is important to recognize that assessments of achievement by norm-referenced tests and by a CBM may be used jointly to operationalize the low achievement and instructional response components of the triangle approach to definition. In Fletcher et al. (2014), no single measure was reasonably accurate in identifying the pool of inadequate responders. However, the norm-referenced assessment of untimed word-reading accuracy and a CBM assessment of oral reading fluency concurred in identifying 90/104 of the pool of inadequate responders; the 14 not identified were all average achievers and likely false positives. This is just one study, but illustrates the type of work that needs to be completed.

Third, the cut points across different measures should be made comparable. This would eliminate one source of variability contributing to low agreement. Different measures utilize different norming populations, which results in difficulties comparing scores across measures. Through efforts to equate scores on different measures, particularly measures of the same construct, some of the error associated with identification decisions may be removed.

Fourth, recognize that agreement across two measures of the same construct is expected, but agreement across measures of different constructs should not be expected. As the TOWRE-CBM example demonstrates, reliability will be enhanced by using two measures of similar constructs. Measures of different constructs measure different theoretical skills and therefore should not identify the same sample of struggling readers.

Fifth, carefully consider the role of growth in assessing LD status. It is clear that a single CBM post-intervention assessment may yield more false positive errors than is desirable and also miss people identified by other measures. However, like all assessments of this sort, such findings are based on specific thresholds and alternatives thresholds should be considered, especially with the psychometric enhancements that are suggested. Adding multiple time points and growth trajectories may help improve the value of CBM assessments, but present research does not show that growth per se increases the accuracy of identifications based on CBM measures. This is of concern because such approaches are proposed as potentially sufficient for identifying LD status (Kovaleski et al. 2013). Slope alone is clearly not adequate to identify LD status, which is why dual-discrepancy methods that incorporate intercept of final status measures are recommended. However, a great deal of work remains to be done to show that assessments of growth add to the information provided by the end-point final status measure. In Fletcher et al. (2014), adding the slope to the final status assessment likely added many adequate achievers to the pool of inadequate responders. Moreover, empirical studies of the value-added information of slopes have not shown that additional information is obtained. Schatschneider et al. (2008) and Tolar et al. (2014) demonstrated that slopes did not add meaningfully to the prediction of outcomes once the final status was entered into the prediction equation. Far from being isolated or atypical examples, these results are exactly what is expected to be found in situations where performance at time 1 is highly restricted due to selection or screening, with the result that fan-spread growth is observed and slopes and final status are highly correlated. When the final status is known following intervention, how much growth the student demonstrated is largely irrelevant. The important issue is whether the cut points applied have been validated. Is there evidence that students who perform below a given level are likely to experience continued academic difficulty?

These problems may reflect the need to study approaches that rely on an index of growth, especially in terms of issues related to cut points, measures, number of time points, and methods for computing slope. In the absence of compelling research, growth-specific approaches to LD identification should not be recommended as assessments of low achievement or as the only criterion for assessing low achievement. Instead, the authors recommend that the low achievement and instructional response components of determinations of LD status be kept distinct at a conceptual level, although they are certainly correlated. It makes great sense to use assessments of growth as part of the determination of instructional response and, in particular, to use this information to modify instruction. But it is uncertain whether dual-discrepancy methods relying on CBM data only are adequately reliable for determining LD status. Moreover, CBM methods for assessing complex integrative skills like reading comprehension or written expression simply are not available. These domains are difficult to assess reliably without a longer assessment and a lot of items.

Finally, the field may benefit from movement towards an approach that considers LD identification in terms of probability or risk. In light of the inherent psychometric difficulties related to the LD identification process, Macmann et al. (1989) suggested that:

> classification systems should at least be based on a coherent psychology of helping…. Although there is no shortage of children who experience problems in adjustment and the acquisition of essential skills, assessments of the characteristics of these children are important to the extent that contributions are made to the design and evaluation of meaningful interventions. (pp. 145–146)

Through this lens, assessment is understood as an ongoing process, incorporating multiple indicators tied directly to treatment outcomes. As assessment and educational placement become more fluid, highly sensitive cut points and confidence intervals could be incorporated to ensure very few students who require additional instruction do not receive it. Ultimately, the authors feel that the field would be well served to move towards an LD identification process that is not focused on issues of classification and entitlement, but instead focuses on identifying academic risk so that appropriate instruction can be provided (Table 1).

Table 1 Implications for practice

1. *Learning disabilities (LD) are dimensional*. The identification process requires a categorical decision. However, the imposition of a categorical structure is inherently arbitrary and students close to the cut point on the separating attribute will be highly similar.
2. *All actuarial methods demonstrate limited reliability*. Actuarial methods apply a specific cut point to continuous data to create categorical groups. However, this results in instability in group membership because of measurement error. This is true for methods based on response to intervention (RTI), as well as alternative methods.
3. *Methods based on RTI demonstrate good validity*. Although unreliable at the individual level, methods based on RTI demonstrate good validity. Adequate and inadequate responders can be differentiated on many attributes, including academic achievement, cognitive performance, subsequent treatment response, and brain functioning.
4. *Cognitive discrepancy models demonstrate poor validity*. Students that meet and do not meet cognitive discrepancy criteria cannot be differentiated on attributes not utilized to form groups, thus failing a critical test of external validity.
5. *Identification decisions should incorporate multiple academic measures*. Single tests should not be utilized for identification purposes. Incorporating multiple tests may: (a) provide better coverage, (b) assess a full range of component skills within an academic domain, and (c) inform subsequent treatment.
6. *Employ confidence intervals and high thresholds for identification decisions*. Limit the use of strict cut points. Instead, decisions should be based on considerations of a range of plausible scores. High thresholds and the use of confidence intervals improve coverage and provide interventions to students who need them.
7. *Utilize curriculum-based measures (CBM) to assess growth and guide instruction*. CBM data play an important role in ongoing progress monitoring. These data can be used to make instructional decisions about pacing and content and to determine instructional response.
8. *CBM may not be sufficient for LD identification*. The reliability of identification decisions based strictly on CBM may not be adequate. Additionally, CBM (at present) do not adequately assess complex skills like reading comprehension or written expression.
9. *Focus assessment on academic needs*. Assessment processes focused on cognitive functioning do not demonstrate good treatment validity, because the efficacy of cognitive training programs is unproven. Instead, assessment should focus on identifying specific areas of academic weakness, so that intervention can be provided.

References

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39,* 414–431.

Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., & Francis, D. J. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences, 18,* 296–307.

Bereiter, C. (1967). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in the measurement of change*. Madison: U of Wisconsin Press.

Bradley, R., Danielson, L., & Hallahan, D. P. (Eds.). (2002). *Identification of learning disabilities: Research to practice*. Mahwah: Erlbaum.

Brown-Waesche, J. S., Schatschneider, C., Maner, J. K., Ahmed, Y., & Wagner, R. K. (2011). Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities, 44,* 296–307.

Burns, M. K., & Senesac, S. V. (2005). Comparison of dual discrepancy criteria to assess response to intervention. *Journal of School Psychology, 43*(5), 393–406.

Burns, M. K., Scholin, S. E., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment, 28,* 102–114.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford Press.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American journal of Mental Deficiency, 86,* 127–137.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P., & Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development, 80*(1), 77–100.

Consortium for Evidence-Based Early Intervention Practices. (2010). A response to the Learning Disabilities Association of America (LDA) white paper on specific learning disabilities (SLD) identification. www.isbe.state.il.us/spec-ed/LDA_SLD_white_paper_response.pdf. Accessed 16 May 2012.

Ellis, A. W. (1984). The cognitive neuropsychology of developmental (and acquired) dyslexia: A critical survey. *Cognitive Neuropsychology, 2,* 169–205.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: Genetic dissection of a complex cognitive trait. *Neuroscience, 3,* 767–780.

Flanagan, D. P., Ortiz, S. O., & Alphonso, V. C. (Eds.). (2007). *Essentials of cross- battery assessment*. New York: Wiley.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic deficits. *Child Development Perspectives, 3,* 30–37.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 6–23.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York: Guilford Press.

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to intervention. *School Psychology Review, 40,* 2–22.

Fletcher, J. M., Stuebing, K. K., Morris, R. D., & Lyon, G. R. (2012). Classification and definition of learning disabilities: A hybrid model. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 33–50). New York: Guilford Press.

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Miciak, J., Francis, D. J., & Denton, C. A. (2014). Agreement and coverage of indicators of response to intervention: A multi-method comparison and simulation. *Topics in language disorders, 34*(1), 74.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3–17.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38,* 98–110.

Fuchs, D., & Deshler, D. K. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 20,* 129–136.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13,* 204–219.

Geary, D. C., Hoard, M. K., Byrd-Craven, M., Nugent, L., & Chattavee, N. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1359.

Grigorenko, E. L. (2001). Developmental dyslexia: An update on genes, brains and environments. *Journal of Child Psychology and Psychiatry, 42,* 91–125.

Grigorenko, E. L. (2005). A conservative meta-analysis of linkage and linkage-association studies of developmental dyslexia. *Scientific Studies of Reading, 9,* 285–316.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Hale, J. B., Alfonso, V., Berninger, B., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33*(3), 223–236.

Hoskyn, M., & Swanson, H. L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *The School Psychology Review, 29,* 102–119.

Jorm, A. F., share, D. L., Matthews, M., & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology, 27,* 45–54.

Kearns, D. M., & Fuchs, D. (2013). Does cognitively focused instruction improve the academic performance of low-achieving students? *Exceptional Children, 79,* 263–290.

Kovaleski, J. F., VanDerHeyden, A. M., & Shapiro, E. S. (2013). *The RTI approach to evaluating learning disabilities*. New York: Guilford Press.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19.

Macmann, G. M., & Barnett, D. W. (1985). Discrepancy score analysis: A computer simulation of classification stability. *Journal of Psychoeducational Assessment, 4,* 363–375.

Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of classification agreement. *The Journal of Special Education, 23*(2), 127–149.

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. A., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40,* 148–182.

Miciak, J., Stuebing, K. K., Vaughn, S., Roberts, G., Barth, A. E., & Fletcher, J. M. (2014a). Cognitive attributes of adequate and inadequate responders to reading intervention in middle school. *School Psychology Review, 43*(4).

Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014b). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly*.

Molfese, P. J., Fletcher, J. M., & Denton, C. A. (2013). Adequate versus inadequate response to reading intervention: An event-related potentials assessment. *Developmental Neuropsychology, 38*(8), 534–549.

Morris, R. D., & Fletcher, J. M. (1988). Classification in neuropsychology: A theoretical framework and research paradigm. *Journal of Clinical and Experimental Neuropscyhology, 10,* 640–658.

Morris, R., Lovett, M. W., Wolf, M., Sevcik, R., Steinbach, K., Frijters, J., & Shapiro, M. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities, 45*(2), 99–127.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York: Wiley.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system*. Chicago: Riverside Publishing.

Nelson, R. J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice, 18,* 255–267.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105–119.

Plomin R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131,* 592–617.

Reschly, D. J., & Tily, W. D. (1999). Reform trends and system design alternatives. In D. Reschly, W. Tilly, & J. Grimes (Eds.), *Special education in transition* (pp. 19–48). Longmont: Sopris West.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Ready or not? Or watch- them- fail. *School Psychology Quarterly, 24,* 130–145.

Rezaie, R., Simos, P., Fletcher, J., Cirino, P., Vaughn, S., & Papanicolaou, A. C. (2011). Temporo-parietal brain activity as a longitudinal predictor of response to educational interventions among middle school struggling readers. *Journal of the International Neuropsychological Society, 17,* 875–885.

Rodgers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology, 53,* 369–373.

Ross, A. D. (1976). *Psychological aspects of learning disabilities and reading disorders*. New York: McGraw-Hill.

Rutter, M. (1982). Syndromes attributed to "minimal brain dysfunction" in childhood. *American Journal of Psychiatry, 139,* 21–33.

Satz, P., & Fletcher, J. M. (1980). Minimal brain dysfunctions: An appraisal of research concepts and methods. In H. Rie & E. Rie (Eds.), *Handbook of minimal brain dysfunctions: A critical view* (pp. 669–715). New York: Wiley.

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences, 18,* 308–315.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326,* 145–150.

Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25,* 618–629.

Silva, P. A., McGee, R., & Williams, S. (1985). Some characteristics of 9-year-old boys with general reading backwardness or specific reading retardation. *Journal of Child Psychology and Psychiatry, 26,* 407–421.

Simos, P. G., Rezaie, R., Fletcher, J. M., & Papanicolaou, A. C. (2013). Time constrained functional connectivity analysis of cortical networks underlying phonological decoding in typically developing school-aged children: A magnetoencephalography study. *Brain and Language, 125*(2), 156–64.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86,* 24–53.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39,* 469–518.

Stuebing, K. K., Barth, A. E., Molfese, P. J., Weiss, B., & Fletcher, J. M. (2009). IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children, 76,* 31–51.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41,* 3–21.

Tanaka, H., Black, J., Hulme, C., Leanne, S., Kesler, S., Whitfield, G., Reiss, A. L., Gabrieli, J. D., & Hoeft, F. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science, 22,* 1442–1451.

Taylor, H. G., Satz, P., & Friel, J. (1979). Developmental dyslexia in relation to other childhood reading disorders: Significance and clinical validity. *Reading Research Quarterly, 15,* 84–101.

Tolar, T. D., Barth, A. E., Fletcher, J. M., Francis, D. J., & Vaughn, S. (2014). Predicting reading outcomes with progress monitoring slopes among middle grade students. *Learning and Individual Differences, 30,* 46–57.

Torgesen, J., Wagner, R., & Rashotte, C. (1999). *Test of word reading efficiency*. Austin: Pro-Ed.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult to remediate and readily remediated poor readers: More evidence against the IQ Achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33,* 223–238.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39,* 157–169.

Wechsler, D. (2003). *Wechsler intelligence scale for children, fourth edition: Technical manual*. San Antonio: The Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca: Riverside.

Contextual Influences and Response to Intervention

Amy L. Reschly and Melissa Coolong-Chaffin

Theories of development have increasingly become more nuanced, integrative, and sophisticated. The classic nature versus nurture debate has been supplanted by a biopsychosocial view of development that includes complex interactions among individuals and environments across the life span (Sameroff 2010). Narrow explanations focused on either nature or nurture have fallen by the wayside in developmental theory; thus, similarly narrow explanations of students' school performance (i.e., those that attribute success or failure solely to parents, schools, or students themselves) must also be dispensed with. Genetics and complex biological–social interactions are not, at this point, easily translated to applied work with students. However, despite general acknowledgment of the importance of social ecology for understanding and informing assessment and intervention processes (Gutkin 2009; Reschly and Christenson 2012a), current practices in the field of education leave much to be desired in this regard.

This chapter was revised and updated from the chapter by Reschly, Chaffin, Christenson, and Gutkin (2007), *Contextual influences and RTI: Critical issues and strategies*, that appeared in the first edition of this volume.

A. L. Reschly (✉)
University of Georgia, Athens, USA
e-mail: reschly@uga.edu

M. Coolong-Chaffin
University of Wisconsin—Eau Claire, Eau Claire, USA

Ecological Systems Theory

In the classic model proposed by Urie Bronfenbrenner (1977, 1992), individuals are understood as developing within a context. Bronfenbrenner further differentiated context into several levels, or systems. These systems are often depicted as a series of embedded circles or as Russian stacking dolls (See Fig. 1). The systems, termed micro-, meso-, exo- and macrosystems, range from proximal to distal in terms of influence. The microsystem consists of those individuals and settings in which a child has direct contact, such as homes, schools, and churches. The mesosystem consists of the interactions among the microsystems, such as the relationship between a student's home and school. The exosystem refers to settings in which the child does not have direct contact but which may affect the child, such as the parent's workplace. The macrosystem is the broad social blueprint in which the other systems are embedded, and may include social norms, values, government, and so forth. Proximal systems have greater effects on development, whereas the effects of distal systems are less direct in terms of influence on student outcomes (Downer and Meyers 2010). The final system, chronosystem, refers to the dynamic interaction among the other systems over time.

There are several rules or principles that govern these systems of development. These are described as follows:

- *Individuals and systems exert reciprocal influence over time*. Development is a process of

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_26

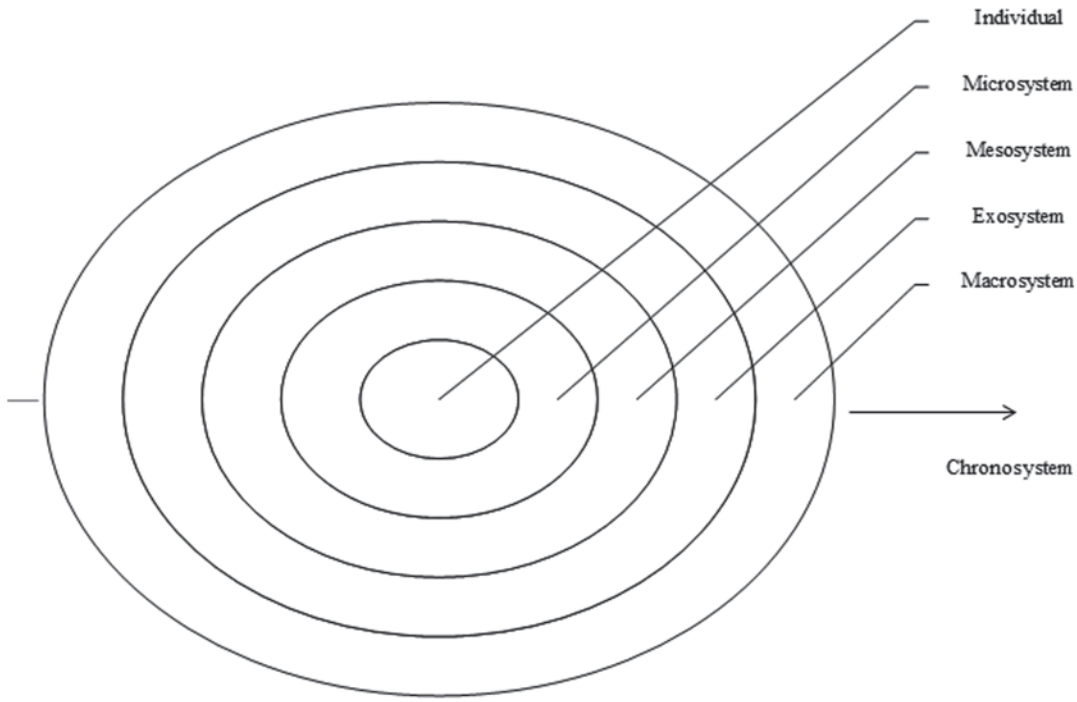


Fig. 1 Bronfenbrenner's ecological systems theory

adaptation between individuals and the contexts in which they are embedded. In addition, as children age, they interact directly with more systems; as a result, greater complexity is needed to understand adolescent development and performance for example.

- *The principles of multifinality and equifinality are necessary to understand the relationship between development and outcomes over several years*. Multifinality refers to the idea that similar initial conditions may result in different outcomes, while equifinality means that different initial conditions may lead to the same outcome (Christenson et al. 1986). Recall that development is a process of ongoing adaptation to the environment and interactions between individuals and systems. The same event may affect individuals differently. Similarly, there may be several developmental pathways to a particular outcome, such as high school graduation.
- *Nonsummativity refers to the idea that the components of a system are greater than the sum of individual parts*. Interactions among the parts result in something greater than each part taken in isolation (Christenson et al. 1986). Pianta and Walsh (1996) extended this concept specifically to the relationships that develop between systems and socializing agents within systems. These relationships are part of the social system that either enhance or thwart students' learning and performance. Discontinuity between systems—particularly home and school—introduces increased risk for students and may be a target of intervention efforts (Pianta and Walsh 1996).
- *Circular causality* refers to the notion that every action within a system is also a reaction (Christenson et al. 1986). Changes over time are not necessarily linear. Further, changes in one system may have a ripple effect to other systems. For example, a change in a student's home environment, such as an illness or moving, may affect peer relationships or behavior at school.

Systems Theory and Response to Intervention

> The most important change for assessment practices has yet to come.Christenson and Ysseldyke (1989, p. 410)

Education in general and school psychology in particular have long been plagued with narrow, simplistic explanations of students' successes and difficulties. These narrow views are reflected in similarly narrow assessment practices. Interestingly, almost all textbook definitions of assessment note that it should be relevant to instruction and intervention (Christenson and Ysseldyke 1989). Indeed, the link to intervention is widely considered paramount to the assessment enterprise (Ysseldyke and Christenson 2002). Unfortunately, traditional, widely used approaches to student assessment are focused on "within-child" explanations of student difficulties (Christenson and Ysseldyke 1989; Sheridan and Gutkin 2000), from which only a tenuous link to relevant targets of intervention may be garnered. Many other shortcomings of assessment practices have been noted, including assessment that is conducted for the purpose of special education eligibility and classification, rather than intervention development and implementation (Reschly 2008); focus on remediation rather than prevention of student difficulties (Reschly and Bergstrom 2009; Shinn and McConnell 1994); too little time spent improving general education (Shinn and McConnell 1994); and the lack of attention to important variables, such as instruction; classroom, school and home environments; rate of learning new material; parent and teacher behavior management; and student engagement and motivation, among others (e.g., Christenson and Anderson 2002; Christenson and Ysseldyke 1989; Gutkin 2009; Reschly 1988; Sheridan and Gutkin 2000).

For many scholars and practitioners, response to intervention (RTI) represents an opportunity to address many of the previously named shortcomings of traditional "refer–test–place" assessment and intervention practices. As described elsewhere in this volume, RTI "is a data-based process to establish, implement, and evaluate interventions that are designed to improve human services outcomes" (Reschly and Bergstrom 2009, p. 434). RTI is inherently intervention-oriented and grounded in progress monitoring, data utilization for decision-making, and evidence-based practices. Most importantly, RTI necessitates a shift from within-child conceptualizations of student difficulties to thinking about learning and behavior within the broad educational environment, which includes both homes and schools. Thus, RTI represents an opportunity to finally reconceptualize assessment and intervention within an ecological systems theory framework. Indeed, the integration of ecological systems theory with RTI is necessary to fully realize the potential of RTI reforms for improving student outcomes. In the following sections, a model for the integration of ecological systems theory and student assessment and intervention is described. Examples of assessment tools and evidence-based interventions that illustrate these principles are also presented.

Ysseldyke and Christenson Model of Student Learning

Drawing from ecological systems theory, Ysseldyke and Christenson (2002) proposed a model to conceptualize student learning within the contexts of school and home (i.e., the contexts where learning occurs), as well as the mutual support or connection between these two primary contexts (Bronfenbrenner's mesosystem; Fig. 1). The model focuses on alterable, relevant variables, or those that have been found to be empirically associated with student achievement and are amenable to effects of intervention. Ysseldyke and Christenson (2002) identified 23 variables related to student learning organized into three categories: instructional support, home support, and home–school support (Fig. 2). The model is intended to facilitate the assessment-to-intervention link and ensure that educational professionals consider the individual nature of student development in context (i.e., person–environment fit).

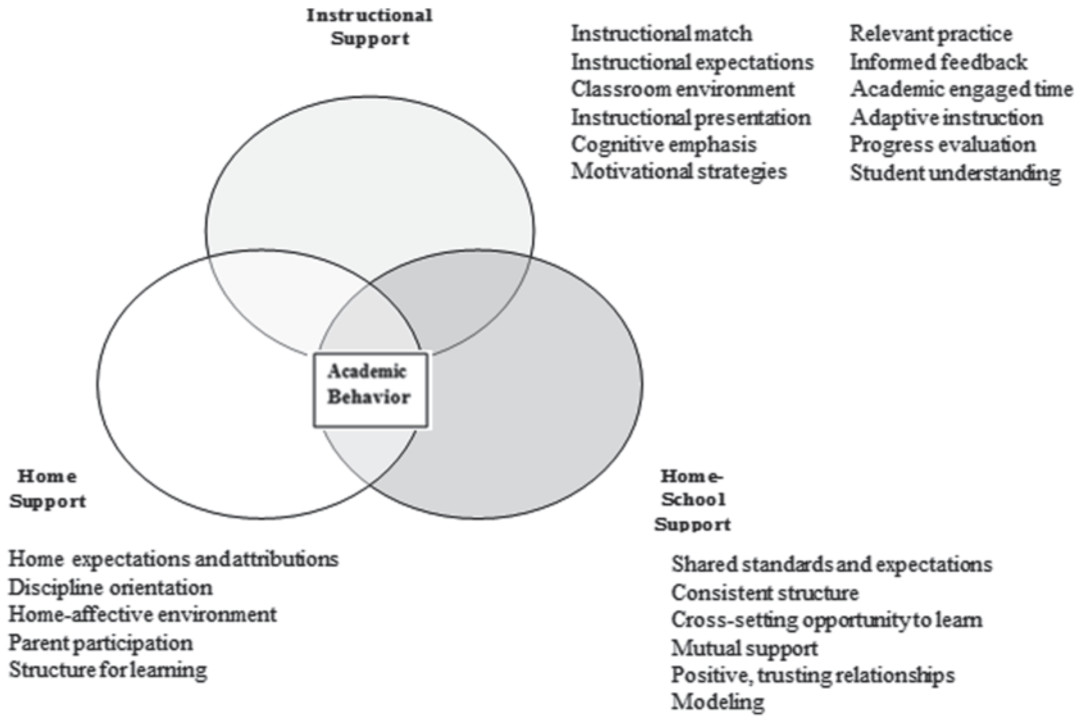


Fig. 2 Ysseldyke and Christneson's contextual model of student learning

Implications for Assessment and Intervention Ecological systems theory and RTI have significant implications for how students' successes and difficulties are conceptualized, data are collected, and design and implement interventions are made (see Table 1). Clearly, if students are understood as developing, learning, and behaving within contexts, then assessment and intervention must account for these contexts. In addition, given that individuals and environments change with time, assessments, too, must account for the ongoing nature of development. Furthermore, outcomes of interest—such as passing grade-level and high school exit exams or dropping out of school——are complex, with multiple pathways and determinants. Thus, examining students and their environments systemically, rather than as single variables in isolation is a must.

The shift in focus from *where* to teach (i.e., general or special education) to *how* and *what* to teach requires that professionals ask different questions (e.g., examining home and school support for learning, opportunities to learn, antecedents and responses to various strategies), utilize different tools (e.g., those that allow for frequent progress monitoring relative to specific skills and goals), and consider different intervention targets (e.g., manipulation of environment to determine effect on student progress). In addition, the integration of ecological systems theory and RTI necessitates changing roles for stakeholders, particularly families. For example, families may be offered the opportunity to be informed and involved at the first indication of a problem or concern and actively partner in the assessment and intervention process. School professionals spend less time in traditional assessment practices and more time in consultation, screening, direct intervention, and program consultation, activities that are well suited to school–family partnering efforts.

In the next sections, examples of assessment tools and intervention strategies and programs that may be used to account for and address critical contextual influences and settings within an

Table 1 Implications for practice

Shift in focus from simplistic, within-child conceptualizations of student learning and behavior to the broad educational context
Recognition that learning environments for students consists of school, home, and home–school components[a]
Assessment and intervention practices account for the broad learning environment (e.g., home and school support, opportunities to learn) and ongoing nature of development (e.g., progress monitoring)
Assessments are (a) linked to the actual contexts and behaviors students are asked to make, (b) used to inform intervention strategies and decision-making, and (c) integrated across a variety of sources, contexts, and methods
Changes to assessment and intervention practices provide an opportunity and means of engaging with families at the first sign of student difficulty
Changes in roles for educational personnel result in less time spent in traditional assessment activities and more time in consultation, direct intervention, and program evaluation

[a] From Ysseldyke and Christenson (2002)

RTI model are described. The tools and programs highlighted in this chapter are not exhaustive, nor does including a tool or program ensure that one has appropriately accounted for important, relevant aspects of students' contexts. Rather, ecological systems theory provides a framework for integrating information across contexts, settings, and interactions. Thus, examining isolated contextual variables, such as how families supervise homework completion or pace of instruction in a student's classroom, is not sufficient to ecologically valid assessment and intervention.

Tools for Ecological Assessment in RTI

The challenge for educators is to conduct assessments that take into account the multiple contexts in which the child is learning as well as the interactions between these contexts and to use that information to make decisions about instructional programming for students. The goal of assessment in an RTI system is not simply to determine whether or not a student qualifies for special education services, rather the assessment process helps practitioners pinpoint what in the environment brings out the best response from the targeted student. Assessment is not a finite step on the road to eligibility, rather it is an ongoing process through which the most appropriate intervention for the student's specific problem is identified, implemented, and its effectiveness evaluated.

To maximize ecological validity, assessment practices should closely approximate natural classroom practices in terms of the context in which the assessment takes place, the materials used, and the responses the student is required to make (Dean et al. 2006). In addition, by definition, assessments should help educators make important decisions about students—whether those decisions relate to resource allocation or instructional and intervention strategies (Salvia et al. 2013). In light of these characteristics, it is proposed that the following considerations should be made when evaluating the extent to which assessment practices are ecological: Assessments should consider the child within and across *natural contexts*, using *natural tasks* which require *natural responses* from the students. Information collected should *directly inform instruction*. Lastly, information from a *variety of sources*, gathered in a *variety of contexts*, through a *variety of methods*, should be considered in an *integrated manner*.

Several commonly used assessment tools and/or approaches that can be used in conjunction as part of an ecologically valid assessment are described. The purpose here is to highlight a sampling of assessment tools and approaches that may be used. The reader is referred elsewhere in this volume for more detailed information about different approaches. Chapters are noted where appropriate.

RIOT/ICEL Matrix A helpful framework for organizing assessment strategies across domains is the RIOT/ICEL matrix developed by the Heartland Area Education Agency 11 (as cited in Christ 2008). Figure 3 represents the adaptation of this matrix. RIOT reminds educators of the importance of a multimethod approach incorporating

	R REVIEW	I INTERVIEW	O OBSERVE	T TEST
I INSTRUCTION				
C CURRICULUM				
E ENVIRONMENT	HOME SCHOOL HOME-SCHOOL			
L LEARNER				

Fig. 3 RIOT/ICEL matrix. (Adapted from Christ (2008))

records reviews, interviews, and observations, and testing (i.e., RIOT), while ICEL prompts us to consider multiple domains (i.e., instruction, curriculum, environment, and finally the learner). In addition, information is gathered from multiple sources through each method and in each domain (Christ 2008). To capture Ysseldyke and Christenson's (2002) conceptualization of the total learning environment (i.e., instructional support, home support, and home–school support), the authors modified the matrix dividing the environment section into its three component parts. Examples of ecological assessment tools or approaches can be organized using the RIOT acronym.

Records Review A thorough review of existing records (e.g., educational, medical, permanent products, etc.) allows practitioners to consider the student's development across multiple contexts, in multiple domains, and across time. Pertinent information collected through records review includes information about health, development, family structure and functioning, educational history, curriculum, and instructional strategies used, past strengths and concerns, etc. The ICEL acronym can be helpful for organizing a review of records.

Interviews Interviews conducted with key stakeholders (e.g., parents, teachers, the student, etc.) yield valuable information from multiple perspectives. Many authors have developed useful interview protocols which allow for exploration of student strengths and needs within and across the various contexts in which the student functions.

Interview protocols developed by Sheridan and Kratochwill (2008) for use within the conjoint behavioral consultation (CBC) process are one example of an ecological approach. CBC applies ecological systems theory and behavioral theory to meet the needs of struggling learners by developing interventions through partnerships across important contexts (e.g., home and school). The series of interviews stress finding shared goals, developing interventions through consensus that can be implemented across contexts, and determining intervention effectiveness based on evidence in multiple settings.

In addition, the Functional Assessment Interview and Student-Directed Functional Assessment Interview Forms developed by O'Neill and colleagues provide an excellent format for exploring problem behaviors, as well as their antecedents and consequences in considerable depth. This information is used to develop interventions that address problem behaviors through prevention and teaching appropriate replacement skills (O'Neill et al. 1997). When considering students from culturally and linguistically diverse backgrounds, questions developed by Rhodes et al. (2005) provide insight into important areas of development (e.g., family history, health and developmental history, educational history, language history, and level of acculturation) that can aid in developing interventions.

Observational Tools Several observational tools exist allowing educators to record real-time information about student behaviors as they relate to factors in the environment. Observation protocols range from simple narrative recordings of antecedents, behaviors, and consequences (i.e., ABC charts), to more complex event recording forms such as the Functional Assessment Observation Form (FAO; O'Neill et al. 1997). Using the FAO, the observer records key behaviors on the customizable form. The FAO includes common predictors such as demands or requests, difficult tasks, and transitions, as well as common functions (i.e., get/obtain and escape/avoid). In addition, the form has space for recording additional predictors and perceived functions. Please see Chap. 22 for more information about functional behavior assessment approaches.

Systematic direct observation forms such as the State-Event Classroom Observation System (SECOS; Saudargas and Lentz 1986) allow for the estimation of percent of time and rates of problem behaviors (e.g., out of seat, calling out, object aggression, etc.) as well as positive behaviors (e.g., school work, raising hand, compliance with directions) through a combination of event recording and momentary time sampling procedures. In addition, teacher behaviors such as approaching the student and communicating approval or disapproval can be recorded. This understanding of behavior in context is essential to developing effective interventions.

Testing Testing within an ecological approach is considered to be broad, including formal and informal measures, standardized, and nonstandardized tests, and well as hypothesis testing approaches.

Curriculum-Based Measurement Curriculum-based measures (CBM) of academic skills are well aligned with ideals of ecological assessment. These tools were originally derived from the curricular materials being used in the classroom, require authentic production responses, and can be conducted quickly and easily by teachers in classroom settings (Deno 1985). For example, CBMs of oral reading fluency (CRM-reading, CBM-R) are useful in informing instruction and can lead to better student achievement (see Stecker et al. 2005 for a review). CBM-R can provide information about student performance at a given point in time relative to peers or other standards, but these measures are also sensitive to small changes in performance, and thus are suitable for ongoing progress monitoring, features critical for use within an RTI model (see Wayman et al. 2007 for a review). Please see Chaps. 12, 17, and 21 for more information about CBM.

Functional Analysis of Academic Behavior From a behavioral perspective, RTI involves a functional rather than a structural explanation for performance deficits (Christ et al. 2005). In contrast to focusing on within-child deficits as an explanation for learning problems (i.e., the structural approach), the functional approach focuses on external, alterable variables affecting the child's performance such as time allotted for instruction, level of difficulty of material, and teacher feedback (Daly et al. 1997). Since the explanatory variables for performance deficits are alterable, they can be manipulated to test various hypotheses about why the problem is occurring. Once a plausible functional explanation is determined, appropriate interventions can be selected based on that function.

Daly et al. (1997) pioneered the use of brief experimental analysis (BEA) for choosing and evaluating academic interventions. Each intervention is designed to test one of the following hypotheses: (1) The child does not want to do the task, (2) The child has not had enough practice to do the task, (3) The child has not had enough help to do the task, (4) The child has not had to do it that way before, (5) The task is too difficult. By manipulating each independent variable successively (i.e., incentive, practice, modeling, rehearsal, and feedback, and task difficulty, respectively), while measuring the same dependent variable (e.g., oral reading fluency), and then replicating the results, the most successful intervention can be chosen for each student. The hypotheses are arranged in ascending order from least intrusive to most intrusive, and when tested in that succession, allow the interventionist to determine the most simple, effective intervention for the student.

Using BEA within an RTI framework allows practitioners to determine not only whether or not a student has "responded to intervention" for special education placement decisions (i.e., where to teach) but also answers the more practical questions of how to teach and what to teach.

An Integrated Approach Ysseldyke and Christenson's *Functional Assessment of Academic Behavior (FAAB*; 2002), takes a comprehensive ecological approach. As an assessment tool FAAB, "maintains a focus on identifying and coordinating instructional, home, and home–school support for the referred student with the express purpose of designing feasible interventions to enhance the student's academic success" (p. vii). The FAAB provides the philosophical framework as well as specific assessment tools for gathering information including reproducible parent, teacher, and student interview and classroom observation forms. Once information is gathered, interventions to address the fit, or lack thereof, between student characteristics and the total instructional environment can be developed. The FAAB takes into account the important influence of home support for learning and the connection between home and school environments whereas many other assessment tools do not.

To summarize, an ecologically valid assessment is one that considers the child within and across *natural contexts*, using *natural tasks* and requiring *natural responses* from the students. Information collected should *directly inform instruction*. Lastly, information from a *variety of sources*, gathered in a *variety of contexts*, through a *variety of methods*, should be considered in an *integrated manner*. A truly ecological assessment requires an understanding of how systems interact with one another to impact student performance. As Sheridan and Gutkin (2000), eloquently illustrated:

> When children experience difficulty learning to read, for example, this "dysfunction" is best understood as the product of multilayered, proximal, distal, and interactive systems. Among these systems are the individual children themselves, educational contexts, prevailing social environments, societal influences, and the interactions among and across all of these systems (p. 486).

Without understanding student difficulties as resulting from the complex interactions within and across systems, and developing comprehensive interventions that address those multiple contextual influences there is little hope of remediating those difficulties.

Ecological Interventions in RTI

Assessment tools and approaches that may be used for ecological assessment of students' behavior and academic performance are inherently linked to relevant intervention targets and/or may be used for progress monitoring and evaluation. Similar to ecological assessment, interventions that are grounded in ecological systems theory consider student performance within and across contexts. For example, positive, behavioral, interventions, and supports (PBIS, described in greater detail in Chap. 33) is a school-wide system of support that aims to establish a culture that supports academic and social success (www.pbis.org). Consistent with ecological systems theory, PBIS conceptualizes the school context broadly to include academic and nonacademic settings (e.g., auditorium, hallways), aims to promote consistency across adults and settings within a school, and utilizes data and assessment practices from natural contexts and tasks that may be used to inform intervention and evaluation.

In this section, however, the focus is on interventions that address both home and school contexts to enhance student outcomes. It has long been recognized that families have an enormous impact on student achievement and behavior; however, meaningful opportunities to involve and collaborate with families in school settings are still too infrequent and unsystematic. Given that most youth spend large portions of their time at home and in school, effects of intervention may be limited when focused on one, rather than both, contexts. Furthermore, there is evidence that collaborative family–school interventions are effective for addressing student difficulties (Carlson and Christenson 2005). Finally, RTI reforms provide the structure and opportunity to partner with families to promote students' success (Reschly and Christenson 2012a).

Conjoint Behavioral Consultation As mentioned previously, CBC marries ecological systems thinking with behavioral theory to develop a process of collaborative problem-solving across school and home environments (Sheridan and Kratochwill 2008). An iteration of Behavioral Consultation (Kratochwill and Bergan 1990), CBC is a model of indirect service delivery that unites school professionals (the consultants) with family and other important caregivers (the consultees) to meet the needs of a student (the client). Following the classic steps of the problem-solving model, the consultant guides the consultee through problem (needs) identification, problem (needs) analysis, plan implementation, and plan evaluation. In keeping with CBC's focus on strengths, "problems" have been reframed as "needs." CBC was developed with the recognition that the issues students face are complex and multifaceted, and thus require coordinated efforts across systems to resolve them. The emphasis is on developing shared goals as well as a sense of shared responsibility (Sheridan and Kratochwill 2008).

Decades of research have demonstrated the effectiveness of CBC in addressing many types of student concerns including externalizing behaviors (Sheridan et al. 2001; Sheridan et al. 2012), internalizing behaviors (Sheridan and Colton 1994), social skills (Sheridan et al. 2012), and academic issues (Weiner et al. 1998). In a review of parent consultation-based interventions, Guli (2005) concluded that CBC had the strongest support among other models for improving student outcomes in school settings.

Incredible Years The incredible years is a set of interrelated, collaborative intervention programs for parents, teachers, and children between the ages of 2 and 8 to (a) promote emotional, social, and academic competence and (b) prevent, reduce, and intervene with emotional problems and aggressive behaviors (Webster-Stratton and Reid 2010). Preschool and early school experiences were targeted because of the difficulties some children experience upon the transition to schooling and the poor outlook for those with early conduct and emotional problems. The parent program aims to strengthen parenting skills and enhance parents' involvement at school. Examples of topics include promoting school readiness, positive discipline, handling misbehavior, child and adult problem-solving, providing and requesting support, and working with teachers. The teacher program is a classroom management intervention given across an academic year that addresses subjects such as proactive teaching (e.g., transitions, warnings, routines), promoting competence, using praise effectively, building relationships, managing behavior and developing plans, and use of incentives. There are both small- and large-group child interventions. The small-group intervention is designed to be used weekly with small groups of children whereas the large-group intervention is a classroom intervention delivered by teachers 2–3 times per week. The child intervention targets learning school rules, positive behaviors, identifying and understanding feelings, problem-solving, managing anger, and being friendly (Webster-Stratton & Reid, 2010). The incredible years programs have been extensively researched and demonstrated positive, lasting effects on child behavior (see Webster-Stratton and Reid 2010, for a review).

Check & Connect Check & connect (C&C) is a comprehensive intervention model designed to promote student engagement and school completion for students who are at risk for poor educational outcomes. Student engagement is alterable, influenced by important contexts, such as home and school, and associated with proximal (e.g., achievement) and distal outcomes (e.g., school completion). Enhancing student engagement is a cornerstone of the most promising dropout prevention programs and high school reform initiatives (Reschly and Christenson 2012b).

There are four main components of C&C: (1) a mentor who works with students and families across years; (2) regularly checking alterable indicators of student performance (e.g., behavior, attendance, work completion, effort); (3) initiating more intensive, individualized interventions at the first signs of disengagement to reestablish and maintain student engagement and participation; and (4) working collaboratively with families, school staff, and community personnel to promote educational success (Christenson and Reschly 2010a; Reschly and Christenson 2006). C&C has been implemented in urban and suburban settings, across school levels, and with students with high-incidence disabilities and those without, with consistently positive results. Compared to controls, C&C students were significantly less likely to drop out of school (Sinclair et al. 1998; Sinclair et al. 2005). Other studies found improvements in attendance and school behavior as well as greater parental support of education (Lehr et al. 2004).

EcoFIT EcoFIT is an ecological approach to family intervention and treatment that was developed out of a series of federally funded intervention trials for students at high risk for problem behavior. The EcoFIT intervention has three primary components: (1) a school-based family resource center (FRC) that addresses school-wide discipline, behavior support, and parenting skills, (2) the family check-up (FCU), which includes an initial interview, ecological assessment, and feedback that addresses motivation, areas of strength and those that require growth; and (3) a menu of intervention and treatment options (Stormshak et al. 2010). EcoFIT has been extensively researched with populations ranging from early childhood to late adolescence, with positive effects in terms of problem behavior and risk (Stormshak et al. 2010). More specifically, studies examining the FCU have found lower amounts of antisocial behavior and substance use (Stormshak et al. 2011), lower rates of growth in family conflict, antisocial behavior, and substance use over time (Van Ryzin et al. 2012), and increases in student self-regulation (Fosco et al. 2013).

Future Research

It has been noted that the theoretical sophistication with ecological systems theory is far ahead that of research and practice (Reschly and Christenson 2012a). Similar statements may be made about RTI. Ellis (2005) suggested that three types of evidence are needed to determine whether an educational reform, such as RTI, has sufficient support to merit widespread dissemination. The first type is a strong theoretical basis. The second type is empirical support in real-world settings. Third, there must be evidence of effectiveness from widespread implementation. There is clearly a strong theoretical rationale for integrating ecological systems theory with RTI assessment and intervention practices. However, more research is needed in the areas of educator preparation and implementation, the reliability and validity of the assessment process, and greater development and implementation of theoretically sound, evidence-based interventions.

RTI requires a significant change in educational practice for teachers, administrators, and school psychologists. In particular, new or updated skills are needed in multisystem assessment and intervention, consultation and collaboration, direct intervention, and program evaluation. Surveys indicate that significant preservice and in-service training may be necessary to ensure educational professionals have the skills necessary to function effectively in RTI. For example, a survey of school psychologists indicated a need for professional development in areas such as progress monitoring in core academic subjects, classroom interventions, functional assessment, and prevention and intervention activities that targeted social and behavioral functioning (Stoiber and Vanderwood 2008). Other data indicate that preservice training for many educators does not prepare them for their roles in RTI (Reschly 2012).

Traditionally, more attention has been paid to evaluating the technical properties of scores derived from assessment tools such as CBM (Fuchs 2004). More research is needed to understand the utility of assessment practices in real-world settings, including how assessment information is used and how student performance is impacted as a result. In addition, some intervention approaches and programs that work across systems to address complex student needs are highlighted; however, it is clear that there are also few examples of such effective programs. In this era of increased accountability, the need for evidenced-based practices surpasses what is currently available. More research is needed on all levels to develop and evaluate the efficacy and effectiveness of such programs.

Understandably, many of the efforts to integrate ecological systems theory with educational practices have focused on school–family partnerships. In recent years, the focus shifted from *why* schools should work with families to *how* and *what works* (Reschly and Christenson 2012a). Much work is needed in terms of understanding the process for effective partnering across home and school and further delineation of evidence-based interventions. In addition, theoretically, congruence between parents and educators is an important target of intervention efforts (Christenson and Reschly 2010b) as high-risk conditions for youth are thought to be those in which students receive different messages from home and school (Pianta and Walsh 1996). However, there is little empirical research detailing associations or effects of congruence on outcomes of interest. The numerous methodological and statistical issues inherent in measuring congruence (Glueck and Reschly 2014) are illustrative of the difficulty of conducting research at levels beyond the microsystem (i.e., home or school effects).

Finally, while many professional organizations for teachers such as the National Education Association (www.nea.org) and the National Parent Teacher Association (www.pta.org) advocate for collaboration across home and school, few teacher education programs include coursework specific to fostering effective partnerships (Chavkin and Williams 1988); educators are very often unprepared for this part of their job (Epstein and Sanders 2006). It often has been suggested that school psychologists may be ideally positioned to forge these connections (Christenson 2003; Sheridan and Gutkin 2000). This idea is mirrored in the Model for Comprehensive and Integrated School Psychological Services, which emphasizes service delivery across systems (National Association of School Psychologists, NASP 2010). Research on the most the effective ways to train educators to collaborate across systems and support these efforts in practice is needed.

Concluding Remarks

As noted in other chapters in this volume, it is clear that sound research and scholarship are still needed to advance the understanding and successful implementation of RTI. It is the contention that ecological systems theory serves as a necessary framework for conceptualizing assessment and intervention in RTI. There are several important implications of ecological systems theory for educational practice. Most importantly, ecologically valid assessment and intervention requires consideration of children within and across contexts, using natural tasks and responses from students, and that assessment information informs instruction and intervention. Furthermore, information must be integrated across these contexts. Ecological systems theory holds great promise for realizing the full potential of RTI to enhance student outcomes.

References

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513–531.

Bronfenbrenner, U. (1992). Ecological systems theory. In R. Vasta (Ed.), *Annals of child development. Six theories of child development: Revised formulations and current issues* (pp. 187–249). London: Jessica Kingsley.

Carlson, C., & Christenson, S. L. (Eds.). (2005). Evidence-based parent and family interventions in school psychology [Special issue]. *School Psychology Quarterly, 20*(4).

Chavkin, N. F., & Williams, D. L. (1988). Critical issues in teacher training for parent involvement. *Educational Horizons, 66,* 87–89.

Christ, T. J. (2008). Best practices in problem analysis. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 159–176). Bethesda: NASP.

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual Confusion within response-to-intervention vernacular: Clarifying meaningful differences. *Communiqué, 34*(3).

Christenson, S. (2003). The family-school partnership: An opportunity to promote the learning and competence of all students. *School Psychology Quarterly, 18,* 454–482.

Christenson, S. L., & Anderson, A. R. (2002). Commentary: The centrality of the learning context for students' academic enabler skills. *School Psychology Review, 31,* 378–393.

Christenson, S. L., & Reschly, A. L. (2010a). Check & connect: Enhancing school completion through student engagement. In E. Doll & J. Charvat (Eds.), *Handbook of prevention science*. Lawrence Erlbaum Associates, Inc.

Christenson, S. L., & Reschly, A. L. (Eds.). (2010b). *Handbook of school-family partnerships* (pp. 327–348). New York: Routledge, Taylor and Francis.

Christenson, S. L., & Ysseldyke, J. E. (1989). Assessing student performance: An important change is needed. *Journal of School Psychology, 27,* 409–425.

Christenson, S., Abery, B., & Weinberg, R. A. (1986). An alternative model for the delivery of psychology in the school community. In S. N. Elliott & J. C. Witt (Eds.), *The delivery of psychological services in schools: Concepts, processes, and issues* (pp. 349–391). Hillsdale: Lawrence Erlbaum.

Daly, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26,* 554–575.

Dean, V. J., Burns, M. K., Grialou, T., & Varro, P. (2006) Comparison of ecological validity of learning disabilities diagnostic models. *Psychology in the Schools, 43,* 157–168.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Downer, J. T., & Myers, S. S. (2010). Application of a developmental/ecological model to family-school partnerships. In S. L. Christenson & A. L. Reschly (Eds.), *Handbook of school–family partnerships* (pp. 3–29). New York: Routledge.

Ellis, A. K. (2005). *Research on educational innovations* (4th ed.). Larchmont: Eye on Education.

Epstein, J. L., & Sanders, M. G. (2006). Prospects for change: preparing educators for school, family, and community partnerships. *Peabody Journal of Education*, *81*, 81–120.

Fosco, G. M., Frank, J. L., Stormshak, E. A., & Dishion, T. J. (2013). Opening the "black box": Family check-up intervention effects on self-regulation that prevents growth in problem behavior and substance use. *Journal of School Psychology, 51,* 455–468.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33,* 188–192.

Glueck, C., & Reschly, A. L. (2014). Examining congruence within school-family partnerships: Definition, importance, and current measurement approaches. *Psychology in the Schools, 51,* 296–315.

Guli, L. A. (2005). Evidence-based parent consultation with school-related outcomes. *School Psychology Quarterly, 20,* 455–472.

Gutkin, T. B. (2009). Ecological school psychology: A personal opinion and plea for change. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 463–496). Hoboken: Wiley.

Kratochwill, T. R., & Bergan, J. R. (1990). *Behavioral consultation in applied settings: An individual guide*. New York: Plenum Press.

Lehr, C. A., Sinclair, M. F., & Christenson, S. L. (2004). Addressing student engagement and truancy prevention during the elementary years: A replication study of the Check & Connect model. *Journal of Education for Students Placed At-Risk, 9,* 279–301.

National Association of School Psychologists. (2010). *Model for comprehensive and integrated school psychological services*. http://www.nasponline.org/standards/2010standards/2_PracticeModel.pdf. Accessed 25 Feb 2015.

O'Neill, R. E., Horner, R. H., Albin, R. W., Sprague, J. R., Storey, K., & Newton, J. S. (1997). *Functional assessment and program development for problem behavior: A practical handbook*. Belmont: Brooks/Cole.

Pianta, R., & Walsh, D. B. (1996). *High-risk children in schools: Constructing sustaining relationships*. New York: Routledge.

Reschly, D. J. (1988). Special education reform: School psychology revolution. *School Psychology Review, 17,* 459–475.

Reschly, D. J. (2008). Paradigm shift and beyond. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 3–15). Bethesda: National Association of School Psychologists.

Reschly, D. J. (2012). *Teacher preparation and response to intervention*. Washington, DC: American Institutes for Research, National Comprehensive Center on Teacher Quality.

Reschly, D. J., & Bergstrom, M. K. (2009). Response to intervention. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 434–460). New York: Wiley.

Reschly, A., & Christenson, S. L. (2006). Promoting School Completion. In G. Bear & K. Minke (Eds.), *Children's Needs III: Understanding and addressing the developmental needs of children* (pp. 103–113). Bethesda: National Association of School Psychologists.

Reschly, A. L., & Christenson, S. L. (2012a). Moving from 'context matters' to engaged partnerships with families. *Journal of Educational & Psychological Consultation, 22,* 62–78.

Reschly, A. L., & Christenson, S. L. (2012b). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement*. New York: Springer.

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. New York: Guilford Press.

Salvia, J., Ysseldyke, J., & Bolt, S. (2013). *Assessment for special and inclusive education* (12th ed.). Belmont: Wadsworth.

Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development, 81,* 6–22.

Saudargas, R. A., & Lentz, F. E., Jr. (1986). Estimating percent of time and rate via direct observation: A Suggested observational procedure and format. *School Psychology Review, 15,* 36–48.

Sheridan, S. M., & Colton, D. L. (1994). Conjoint behavioral consultation: A review and case Study. *Journal of educational and Psychological Consultation, 5,* 211–228.

Sheridan, S., & Gutkin, T. (2000). The ecology of school psychology: Examining and changing our paradigm in the 21st century. *School Psychology Review, 29,* 485–502.

Sheridan, S. M., & Kratochwill, T. R. (2008). *Conjoint behavioral consultation: Promoting family-school connections and interventions*. New York: Plenum Press.

Sheridan, S. M., Eagle, J. W., Cowan, R. J., & Mickelson, W. (2001). The effects of conjoint behavioral consultation: Effects of a four-year investigation. *Journal of School Psychology, 39,* 361–385.

Sheridan, S. M., Bovaird, J. A., Glover, T. A., Garbacz, A., Witte, A., & Kwon, K. (2012). A randomized trial examining the effects of conjoint behavioral consultation and the mediating role of the parent-teacher relationship. *School Psychology Review, 41,* 23–46.

Shinn, M. R., & McConnell, S. (1994). Improving general education instruction: Relevance to school psychologists. *School Psychology Review, 23,* 351–371.

Sinclair, M. F., Christenson, S. L., Evelo, D., & Hurley, C. (1998). Dropout prevention for high-risk youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children, 65,* 7–21.

Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children, 71,* 465–482.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795–819.

Stoiber, K. C., & Vanderwood, M. L. (2008). Traditional assessment, consultation, and intervention practices: Urban school psychologists' use, importance, and competence ratings. *Journal of Educational and Psychological Consultation, 18,* 264–292.

Stormshak, E., Dishion, T., & Falkenstein, C. (2010). Family-centered, school-based mental health strategies to reduce student behavioral, emotional, and academic risk. In S. L. Christenson & A. L. Reschly (Eds.), *Handbook of school-family partnerships* (pp. 228–245). New York: Routledge.

Stormshak, E. A., Connell, A. M., Veronneau, M.-H., Myers, M. W., Dishion, T. J., et al. (2011). An ecological approach to promoting early adolescent mental health and social adaptation: Family-centered intervention in public middle schools. *Child Development, 82,* 209–225.

Van Ryzin, M. J., Stormshak, E. A., & Dishion, T. J. (2012). Engaging parents in the Family Check-Up in middle school: Longitudinal effects on family conflict and problem behavior through the high school transition. *Journal of Adolescent Health, 50,* 627–633.

Wayman, M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41,* 85–120.

Webster-Stratton, C., & Reid, M. J. (2010). A school-family partnership: Addressing multiple risk factors to improve school readiness and prevent conduct problems in young children. In S. L. Christenson & A. L. Reschly (Eds.), *Handbook of school–family partnerships* (pp. 204–227). New York: Routledge.

Weiner, R., Sheridan, S. M., & Jenson, W. R. (1998). Effects of conjoint behavioral consultation and a structured homework program on math completion and accuracy in junior high students. *School Psychology Quarterly, 13,* 281–309.

Ysseldyke, J. E., & Christenson, S. L. (2002). *FAAB: Functional assessment of academic behavior: Creating successful learning environments*. Longmont: Sopris West.

Using Single-Case Design in a Response to Intervention Model

T. Chris Riley-Tillman and Daniel M. Maggin

The purpose of this chapter is to describe a response-to-intervention (RTI) evaluation framework for improving the reliability and defensibility of identifying students with academic or behavioral problems that require more intensive supports than the typical student gets. Much of the RTI literature has focused on the use of intervention practices and strategies that are grounded in empirical evidence. The judicious use of research to guide intervention selection is certainly an important part of developing and implementing an effective RTI framework. However, the identification and adoption of evidence-based practices alone is insufficient for appropriately implementing a RTI approach. Among those features that are critical but often overlooked, is the monitoring and analysis system under which the practice is being used. That is, the identification of students that are not successfully responding to various academic and behavioral methods requires a systematic approach for evaluation and monitoring that allows for efficient and reliable monitoring of student progress and RTI (Christ et al. 2013). The framework advocated for in this chapter is based on single-case research methodology, which is a widely used experimental approach for validating educational, behavioral, and psychological interventions. The methodological principles underlying single-case research make it ideally suited for evaluating intervention effects and making data-based decisions in applied settings such as schools and classrooms (Riley-Tillman et al. 2013). This chapter focuses on the advantages, methods, and additional considerations that are needed to establish a successful formative evaluation approach based on single-case research methodology.

As a brief overview, single-case design is a class of experimental methods designed to provide evidence about the relationship between an intervention and a target behavior (Cooper et al. 2007). Unlike research that uses large groups of participants to then make a statement about some population, single-case design focuses on one case (e.g., a single student). This method of experimental analysis has a long history with thousands of published studies. For this reason, single-case design is ideally suited for schools where one student is often the focus.

While single-case design is an ideal method for those who are interested in experimentation with a single student, it is fair to ask whether such an approach is necessary in schools using an RTI model. In the case of RTI, it is understood that the goal is to link an effective intervention to a child in need. While it would be helpful if some form of assessment could be done to be sure of what interventions will work, limitations in assessment, interventions, as well as the realities of applied intervention work make this impossible. There are simply no diagnostic tools to identify the exact approach for a specific case. Further,

T. C. Riley-Tillman (✉)
University of Missouri, Columbia, MO, USA
e-mail: rileytillmant@missouri.edu

D. M. Maggin
University of Illinois at Chicago, Chicago, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_27

there are no interventions that are universally effective for all children. Finally, even when an appropriate intervention is selected, it cannot be assumed that every interventionist will implement said intervention in an ideal manner. In the end, an effective RTI process not only is charged with selecting a relevant intervention but also then implementing the intervention to see if it works for the child/intervention dyad. This is where single-case design comes into play. Statements about the effectiveness of an intervention are based on the observation of a change in a target behavior post application. It is critical that teams understand exactly how much evidence they have about the relationship between the intervention and the target behavior so that they do not make statements or decisions based on specious beliefs.

Throughout this chapter, methods to determine two levels of understanding the intervention/behavior relationship will be discussed. First, it is critical for a team to be able to say with confidence if a child's behavior has changed post intervention. This is not to say that a specific intervention caused the change, but simply that the target behavior changed. In some other cases, it will become important for a team to also determine if a child responded to a specific intervention. This evidence can be used to support the need for special education (e.g., in the case of a child responding to a highly resources intensive intervention protocol) or to suggest generalizing the intervention into other settings. The extension of the evidence movement from the identification of empirically supported academic and behavioral strategies to a process of formative evaluation requires the use of procedures grounded in research to ensure that accurate and objective conclusions are drawn. Fortunately, the single-case research methodology provides a rigorous, data-driven approach that can be readily adapted for use within applied settings. Before considering the use of single-case research in applied settings, it is necessary to have an understanding of the rationale and procedures that underpin the methodology. As such, the following sections provide an introduction to the experimental single-case research paradigm. This discussion will be followed by a description of a variety of considerations for adapting the experimental framework to applied settings.

Logic of Experimental Single-Case Design

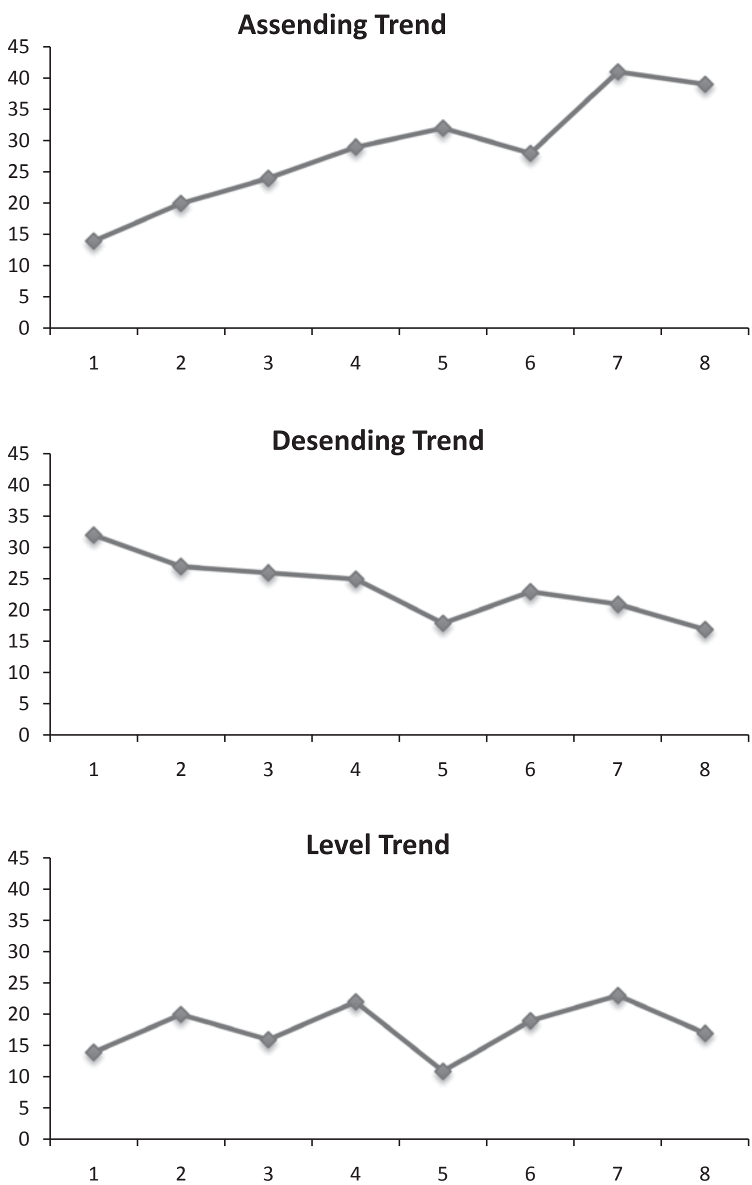
The purpose of all experimental research is to provide a framework for evaluating the direct influence of an intervention, strategy, or practice on some outcome variable of interest (Kazdin 2011). An experimental approach to research facilitates interpretation of the effects that an intervention has on an outcome variable by accounting for potential influence of other plausible explanations. The ability to isolate the effects of an intervention on a behavior or skill of interest allows for more accurate inferences to be drawn about the effectiveness of the strategy. Several experimental frameworks have been developed to address the needs of various disciplines. As such, the single-case research design represents one approach for answering experimental questions. This class of designs is most often used to address questions about the extent to which program components working alone or together impact individual performance. This set of designs is premised on the use of repeated measures in both the absence and presence of a particular strategy (or set of strategies) to determine whether individual responding has changed across conditions. The single-case research paradigm's focus on examining response patterns within individual participants, contrasts with more widely used experimental group-based methods that infer the effects of an intervention to individuals based on group-level responses. These response patterns are obtained through continuous assessment of a targeted behavior or skill over time. As such, data representing the outcome variable of interest is collected on multiple occasions within separate phases. Each phase is differentiated based on the procedural modifications of the intervention being investigated (Kennedy 2005). Procedural modifications can refer to either the introduction of a novel practice or the systematic alteration of an existing practice with the goal of influencing the behavior or skill of concern. The inferences about the effectiveness of an intervention are made through comparisons of the response patterns observed across different conditions for the individual. The strength of these inferences is based on the extent to which the implementation of the design procedures meet established conventions, which are reviewed in the following sections.

Collection of Repeated Measures The collection of repeated measures of the behavior or skill being targeted is the most fundamental requirement of single-case research (Johnston and Pennypacker 2009). That is, the individual's performance on the behavior of concern is continually observed across several occasions. These observations typically occur on a systematic schedule that is adjusted based on the sensitivity of the behavior or skill to various conditions. Whereas classroom behaviors such as disruptiveness or on-task engagement would likely be monitored on a daily schedule, academic skills such as phonemic awareness or vocabulary knowledge might be better suited for a weekly assessment schedule. The resulting data stream is a basic requirement for single-case research because it allows the investigator to examine the pattern and stability of performance both prior to the introduction of the intervention and recurrently while it is in place. The data collected before the procedural modification is initiated provides an assessment of individual functioning without the intervention. Following the introduction of the intervention strategy, the repeated collection of the response variable is continued and the resulting pattern can be compared to those collected in the absence of the intervention. This allows the investigator to determine whether improvements in the target behavior coincide with the implementation of the intervention. As such, the collection of repeated measures over time provides the information needed to make comparisons within the individual research participant.

Establishing Baseline Level of Performance A majority of experimental single-case research designs require the collection of repeated measures of the outcome variable before the introduction of the intervention. As previously mentioned, the continuous measurement of the response variable in the absence of the intervention allows the investigator to determine the initial level of the behavior or skill. Referred to as the *baseline phase,* these measures of the untreated behavior not only provide a descriptive assessment of current behavioral or academic functioning but also serve a predictive function as well (Kazdin 2011). That is, the evaluation of an intervention requires having a reasonable basis for predicting individual performance in the future if the intervention were never to be implemented. For instance, group experimental research involves two (or more) groups of participants in which some groups receive the intervention and some do not. Those not being exposed to the intervention are used as the basis for estimating the level of untreated performance. The single-case approach to experimental research does not typically allow for the estimation of future performance because the ultimate goal is to apply the intervention and evaluate its effects. As such, the data collected for the baseline phase are used to predict what individual performance would have been in the immediate future without intervention. This prediction is based on a projection of the response pattern obtained during baseline into the future. The importance of the predictive utility of future baseline performance within single-case research lies in its subsequent use as a criterion on which to evaluate whether the intervention led to changes in the response variable. Put another way, it can be assumed that, for the intervention to have its intended effect, the pattern obtained on the targeted behavior or skill would need to be different from that obtained from baseline. The quality of the baseline data is therefore critical for providing a sufficient basis for making predictions of future performance and evaluating individual responses to selected interventions.

The utility of a baseline phase for predicting future levels of performance is related to the stability of the data pattern obtained from the collection of repeated measures on the target behavior or skill before the introduction of the intervention (Sidman 1960). A stable data stream refers to a pattern without trend and with only minimal variability. It should be noted that data collected on a particular behavior might reflect both trend and variability but each presents its own set of challenges for prediction of future performance. For instance, trend refers to a consistent or systematic increase or decrease in the observed data over time. Whether the presence of trend in the data is problematic for evaluating the effects of an intervention depends on its direction in relation to the desired change in the behavior or skill. That is, the observed trend might be going either towards or away from the intended therapeutic direction. Figure 1 provides an illustration of the three types of trend that might occur within baseline. Data trends that move in the opposite direction of the anticipated intervention effect do not interfere with evaluation because the therapeutic impact can still be attributed to the intervention if the behavior improves following its introduction. Such is not the case, however, if the baseline data trend is moving towards the hypothesized direction of the treatment. Under these condi-

Fig. 1 Illustrations of three types of trends



tions, the investigator would have difficulty attributing the observed effect to the intervention rather than some other competing variable such as natural maturity or the prior introduction of an undocumented intervention. An additional threat to stable data patterns is variability, which refers to fluctuations in the subject's performance over time. Extreme variability in baseline data impedes making predictions about an individual's future performance by limiting the investigator's ability to anticipate the placement of the next set of data points. As such, data that is characterized by a high degree of variability often do not provide a useful criterion for estimating performance in the immediate future, thereby limiting its value for evaluation. It should be noted, however, that variability in the data is a relative concept and must be considered in relation to the initial level of behavior and the magnitude of expected behavior change following the application of the intervention. Regardless of this caveat, the general rule is that the greater the variability in the data, the more difficult it is to draw conclusions about the effects of the intervention.

Application of the Intervention Strategy The purpose of the single-case research framework is to evaluate the effects of an intervention on individual levels of performance. The basic questions addressed by this set of designs are whether a particular intervention leads to a reliable change in some response variable. Consequently, the single-case research design provides a direct test of some procedural modification on an identified behavior or skill of concern. As previously suggested, procedural modifications might include the application of a novel intervention, the examination of a subset of components from a larger treatment package, or the systematic adaptation of an existing intervention protocol. Each of these approaches to modifying procedural features of the context results in an examination of the conditions for improving individual outcomes for an identified behavior skill. Regardless of the particular modifications being investigated, the single-case research design facilitates the direct evaluation of the intervention through comparison of the data streams collected in baseline and intervention phases. The intervention phase serves the same function as baseline in terms of describing the individual's current level of performance and predicting performance in the immediate future with the added opportunity to test the effectiveness of the intervention. The successful evaluation of an intervention requires the investigator to be able to attribute changes in individual functioning to the selected treatment rather than other variables. Whereas researchers using a group-based approach apply statistical procedures to demonstrate group equivalence or to control for differences between groups on conceptually significant variables, single-case investigators rely on the observed data patterns to evaluate whether changes in the behavior or skill are related to the intervention. It is therefore critical that the obtained data patterns are sufficiently stable so that their descriptive and predictive utility for interpreting intervention effects is maximized. The single-case researcher continually examines the data to determine if enough stable data has been collected to allow for a phase or intervention change. This formative use of single-case data is critical for drawing accurate interpretations from the data (Gast 2010). It further represents a primary strength of single-case research in its ability to facilitate a deeper understanding of the case being investigated.

Verifying Intervention Effects The initial application of the intervention provides preliminary evidence that the treatment is effective for improving the targeted behavior or skill. The term preliminary is used because the observed effect of the intervention represents only a single demonstration that the treatment led to altered patterns of behavior. Even large changes in the response pattern across the initial baseline and intervention phases does not allow for alternative explanations to be ruled out with a single comparison. It is possible that the introduction of the intervention happened to coincide with some other event that was responsible for the observed change in the targeted behavior or skill. Within the logic of experimental single-case research these alternative explanations are ruled out by verifying the original prediction that the baseline level of the behavior would have maintained if the intervention was never implemented. The most straightforward approach for testing the accuracy of the initial prediction that the baseline data series would continue similarly in the immediate future is to simply withdraw the intervention and return to the conditions associated with the original baseline. The prediction statement associated with the initial baseline phase would be successfully verified if the response pattern returned to their original levels. That is, the consistency of individual performance across the two or more baseline phases substantiates the notion that the behavior would have remained at initial levels for the immediate future if left untreated (Cooper et al. 2007). The process of verification provides additional evidence that the intervention was responsible for behavior change rather than some other variable because behavior was shown to change following the implementation of the intervention and to subsequently return to baseline levels after its withdrawal.

The preceding discussion of verification presents a unique problem for behaviors that are not easily reversed (Kennedy 2005). A reversible behavior refers to a targeted behavior that will readily return to baseline levels of performance after the intervention has been withdrawn. Academic skills are a prime example of behaviors that often do not reverse because the individual acquires the ability to perform some task that is not expected to be unlearned following the removal of the intervention. The challenge that irreversibility presents for the logic of verification in single-case research is that the initial prediction associated with baseline performance is not easily verified simply by removing the intervention. Fortunately, there are two approaches available to the single-case researcher that addresses this potential dilemma (Riley-Tillman and Burns 2009). The first is to apply the verification phase as before with the expectation that individual functioning will not return to baseline levels but that the learning pattern will resemble those observed in the initial baseline phase. A return to the original learning trend observed in baseline can be used to verify the prediction that learning would have continued to proceed at the original rate if the intervention had not been implemented. The second and more common approach is to use research designs such as the multiple baseline design that verify the prediction through systematic replication.

Replicating Intervention Effects The ultimate goal of experimental research of any sort is to provide evidence that the intervention is associated with changes in individual responses on some outcome of interest (Kazdin 2011). Evidence refers to the extent to which alternative explanations for the effect of the intervention on the behavior or skill can be ruled out. The presentation of the experimental single-case research paradigm has emphasized the use of prediction and verification to provide empirical support that the original levels of functioning on the targeted behavior are reliably altered through introduction of the intervention (Riley-Tillman et al. 2013). However, the effect of the intervention on individual performance has only been demonstrated once to this point, ultimately raising questions about whether the observed change was actually related to the treatment protocol. Replication of the initial intervention effect is critical for drawing accurate inferences regarding whether the treatment was responsible for observed changes in individual response patterns. Recall that the data streams obtained in each phase provide a basis for describing current levels of performance and making a prediction about functioning in the immediate future. It is necessary therefore to validate the initial intervention effect through another application of the treatment variable, just as it was necessary to verify the baseline prediction that individual performance would have continued along the same trajectory if the intervention was not implemented. Replication of the intervention effect provides a basis for testing the veracity of the behavior pattern obtained with the initial application of the treatment. A replication of the data series provides additional support that the intervention was in fact responsible for the changes in behavior rather than some other variable. That is, multiple demonstrations of the treatment effect increases confidence in the assertion that behavior responded to the intervention by reducing the probability that an alternative explanation is available for behavior change and demonstrating for a second time that the intervention coincided with improvement in the targeted behavior or skill.

Notation of Single-Case Designs The preceding overview provided a description of the four distinct phases of experimental single-case research methodology and their underlying rationale. Specifically, these four phases included, (a) an initial baseline phase in which individual performance is continually assessed without implementation of the intervention, (b) an intervention phase during which an identified treatment is introduced and remains in place to monitor individual responsiveness, (c) a return to baseline conditions accomplished by withdrawing the intervention with the ultimate goal of verifying that individual performance would have maintained if the intervention had not been implemented, and (d) a second application of the intervention with the goal of replicating the data pattern obtained under the initial implementation of the intervention to provide evidence that the treatment was in fact responsible for observed behavior changes. The convention of single-case research is to describe these phases using a notation system that facilitates brief communication and can accommodate a variety of procedural modifications (Cooper et al. 2007). This coding system assigns a letter to each phase with similar conditions allowing for designs to be described with short letter combinations. For instance, a capital letter "A" is used to denote baseline conditions in which data are typically collected in the absence of some procedural modification. As such, both the initial baseline and verification of the baseline would be labeled with an "A" given the four conditions previously described. A capital letter "B" is typically reserved for conditions in which the procedural modification has been used; this applies to both the initial application of the intervention and the replication phase. Taken together, the four phase approach would be denoted as an ABAB design, indicating that repeated measures of the targeted behavior or skill were taken to demonstrate current functioning and predict the immediate future; an intervention was introduced to evaluate whether behavior patterns changed; a second baseline was then used to verify the prediction of individual responding if the intervention had not been implemented; and a final phase was used to replicate the effects of the intervention variable.

This notation system can be extended or truncated based on the arrangement of phases. A truncated version of the system might be used to denote a circumstance in which the research was unable to implement the replication phase resulting in an ABA research design. The single-case research design can also be extended to include an additional procedural modification which would be represented with a capital letter "C." Consider the circumstance in which a researcher augmented the original intervention with an additional modification before returning to baseline, this design would be denoted as an ABC indicating that there was an initial baseline followed by two conditions with unique intervention procedures. The following discussion regarding the adaptation of experimental single-case research methodology for applied settings will rely on this notation system to describe the ordering of particular conditions. This presentation will emphasize some of the strengths and limitations of various design methods for facilitating the evaluation of intervention effects within a response-to-intervention framework.

Adapting the Single-Case Design for Applied Settings

The purpose of the preceding overview was to describe the principles and procedures of the experimental single-case research methodology to provide perspective on its underlying logic for developing empirical evidence. Fortunately, the experimental single-case framework has several features that make it particularly useful for evaluating student responses to selected interventions. For instance, single-case research requires the investigator to continually collect and examine data over time to determine whether the intervention is working as intended. This ongoing routine of gathering, graphing, and scrutinizing data allows school personnel to use the data formatively and provides a transparent, objective, and empirical rationale for making programming decisions. The significance of the single-case process for practice relates to the emphasis on systematic, repeated data collection, establishment of a baseline level of performance, and continual monitoring of intervention effects. Taken together, these elements make for an expedient framework for monitoring student progress that has the further advantage of having been drawn directly from research. As such, the adaptation of the experimental single-case framework for use in applied settings further extends the notion of evidence-based practice from intervention selection to intervention evaluation.

It is reasonable to assume that many school personnel engaged in program planning might question the utility of demonstrating that a particular intervention is responsible for observed changes in the targeted behavior or skill over and above other alternative variables. However, the extension of single-case research into applied settings does not necessarily require an experimental analysis of intervention effects on a selected outcome variable. There might be some instances where school personnel are interested or mandated to determine whether an intervention is functionally related to a behavior or skill of concern but this is not a prerequisite. As such, selection of an appropriate design will depend on the behavior or skill being targeted and the overall purpose of the evaluation (Riley-Tillman and Burns 2009). The following review provides a description of common single-case designs that can be used for a variety of purposes in applied settings. While some of these designs allow for experimental evaluations to be conducted, others do not. The following descriptions will emphasize the rationale and the conditions that are most suitable for each design within a response-to-intervention framework. Please be aware that this presentation is meant to be illustrative rather than comprehensive with several alternative design approaches available for facilitating evaluation.

Non-Experimental Single-Case Research Designs The ultimate goal of a response-to-intervention framework is to provide an empirical basis for making programming decisions for individual students (Brown-Chidsey and Steege 2010). These programming decisions encompass whether a particular student requires additional support that can be provided within the current instructional context of the school or if additional supports need to be provided through expenditures of special education resources. A rigorous response-to-intervention model emphasizes the use of systematic evaluation for adjusting instructional protocols and ultimately facilitating referrals of students to special education. Fortunately, the single-case research methodology is comprised of several distinct phases that can be used individually or together to address the range of evaluation issues confronted within a response-to-intervention framework. The particular evaluation design selected must be appropriately aligned with the type of decision being made. Consequently, the following presentation of non-experimental single-case designs describes an approach for identifying students that might be struggling and subsequently evaluating whether already established secondary interventions are sufficient for remediating these problems.

A Design One purpose of a response-to-intervention framework is to provide an empirical basis for identifying students in need of additional academic or behavioral support (Burns et al. 2012). For this approach to work well, all students must have an opportunity to be evaluated on a range of behaviors or skills that might indicate potential areas of concern for individual students. The collection of a baseline stream of data for each student provides the basis for identifying those that might not be responding to typical instruction for that school (e.g. the use of periodic school-wide assessments aligned with state standards or teachers utilization of progress monitoring tools) (Chafouleas et al. 2013). The baseline data collection for individual students is, in essence, documenting their response to the universal level of support. Thus, when further intervention is not applied, this design can be considered a "running B design" or the continued documentation of the typical intervention (universal level of support) provided in the schools. It is clear that the collection of baseline data alone does not provide sufficient information to infer the effects of universal instruction on student outcomes, but this is not the purpose of monitoring the progress of all students. Rather, baseline data for all students allows school personnel to identify outlying students that might be candidates for additional support. The practice of identifying students based on comparative analyses with other students in the school allows for students exhibiting less intense behavioral or academic skills to be identified more readily. In a sense, this first step is akin to the check engine light in a car that tells you when something is atypical, but not specifically what is wrong.

AB Design Following the identification of students that might require additional academic or behavioral support, school personnel are charged with the task of developing or identifying an intervention that will be used to address the problem. The application of this intervention initiates the second step of the single-case design framework, which is to determine whether implementing the intervention results in changed patterns in the target behavior or skill (Riley-Tillman et al. 2013). From an experimental perspective, the intervention phase tests whether the procedural modifications led to a data stream that deviated from the pattern predicted from baseline observations. Despite providing the ability to compare individual functioning both in the absence and presence of a developed intervention, the AB design limits our ability to determine if the observed change is directly related to the intervention, since this requires the experimental phases of verification and replication previously described. A practical analysis of the AB design, however, suggests that this framework might be optimal for evaluating the effects of secondary interventions in applied settings. That is, there is often no need to rule out alternative explanations for observed intervention effects for students receiving secondary supports that have been systematized in school or classroom settings. As such, the AB design is often more than sufficient for determining whether established programs result in improved outcomes for those students deemed non-responsive to the universal instruction operating in the setting. It should be noted however that despite the appropriateness of the AB single-case design for many cases encountered within a response-to-intervention framework, the approach also represents the minimum condition for evaluating intervention effects. School personnel can only determine if the intervention was associated with the desired changes in behavior or skills if there is documentation of individual performance under both baseline and treatment conditions. This makes the AB design essential for the implementation of a successful response-to-intervention framework.

ABA Design An ABA design indicates that repeated measures of a target behavior or skill were collected during some baseline period followed by the implementation of an intervention that was subsequently withdrawn to return to the original baseline conditions (Burns and Riley-Tillman 2009). The surface limitations of the ABA design are evident in that the return to baseline seems logically out of step with the experimental framework. Specifically, the return to baseline seems not to be warranted regardless of whether the intervention is working or not. That is, the practical advantage of the previously described AB design relates to its ability to provide a framework for evaluating intervention effects without having to withdraw the intervention. This logic does not extend, however, to circumstances in which the intervention is found to be ineffective. Following the evaluation that an intervention is ineffective, it is useful to return to the original baseline conditions to provide further verification of individual functioning in the absence of an intervention. A return to baseline after the student has been found to be non-responsive might reveal that the intervention had a delayed or partial effect on the behavior. A delayed effect would be demonstrated through improvement in the targeted behavior or skill following the removal of the intervention. School personnel might determine that these improvements were related to the original procedural modifications and ultimately decide to reintroduce the intervention. A partial treatment effect might be observed if the behavior deteriorates beyond those observed in the initial baseline phase. Under these conditions, the intervention might have been responsible for sustaining the behavior at the original baseline levels. Regardless of whether the return to baseline verifies the data patterns associated with the initial baseline or not, the use of a second baseline facilitates subsequent evaluations by providing the opportunity to reassess the current level of behavioral functioning and make a renewed prediction about its pattern in the immediate future.

Experimental Single-Case Research Designs The foregoing presentation of non-experimental single-case research designs describes a framework that is useful for identifying and evaluating intervention effects applied to students that are not responsive to universal or secondary supports. These evaluation procedures represent a subclass of single-case research designs that cannot successfully rule out alternative explanations for intervention effectiveness. The designs described in the following sections allow for evidence to be collected about the association between intervention and outcome variable. An experimental evaluation of intervention effects is warranted for students receiving intensive, special education services because it allows for the determination that the selected intervention is in fact associated with treatment outcomes (Riley-Tillman et al. 2013). Without this knowledge, the expediency and durability of intervention effects would likely be comprised.

Withdrawal/Reversal Design The experimental single-case research design that follows the previously described logic most directly is the classic withdrawal design (Kazdin 2011). This is because behavior is continually assessed across each of the experimental phases including baseline, intervention introduction, verification, and replication. As such, the notation for this design is ABAB indicating that there is an initial baseline evaluation of individual performance (A) followed by the application of some intervention (B) which is then withdrawn to return to the original baseline conditions with the goal of verifying initial baseline levels (A) with the effects of the intervention subsequently replicated through reimplementation of the treatment (B). An example of an ABAB design is presented in Fig. 2

The correspondence of the ABAB withdrawal design to the experimental logic of single-case

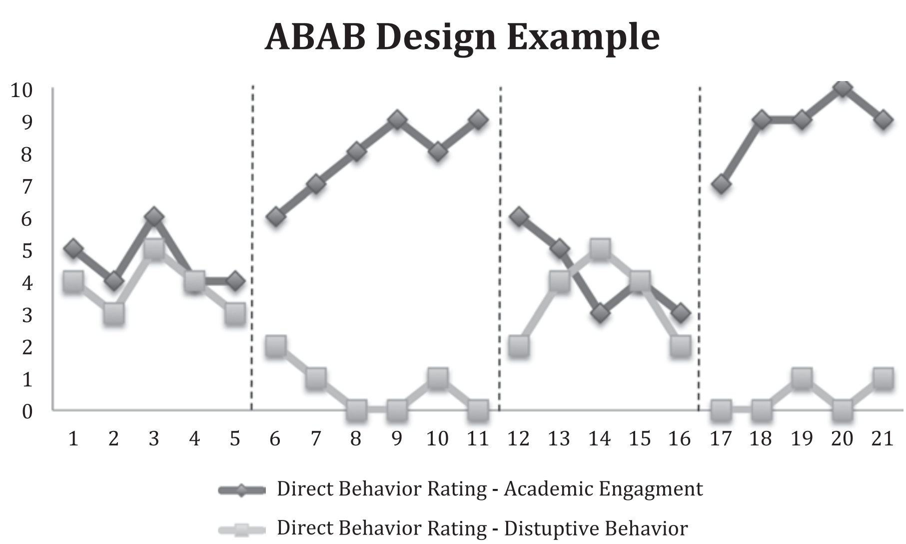


Fig. 2 Example of an ABAB intervention

research has made it among the most widely used approaches for investigating intervention effects in educational contexts (Riley-Tillman and Burns 2009). Despite the internal strength and widespread use of this research design, it has some limitations that led to the development of other experimental frameworks that can be applied if the conditions warrant the use of alternative approaches. Perhaps the most salient limitation of the ABAB withdrawal design is that the intervention has to be removed to verify baseline performance levels. The decision to withdraw an intervention that is seemingly working requires the investigator to weigh both practical and ethical considerations against the potential benefits gained from conducting an experimental analysis. Issues driving these decisions might relate to the severity of the behavior and the anticipated length of the intervention. An additional obstacle to the use of an ABAB design is the nature of the behavior or skill being targeted. Recall from the presentation of the experimental single-case framework that behaviors such as academic skills can pose analytic challenges for verifying original baseline functioning because these behaviors are not expected to readily return to previous levels. In such cases, it is understood that a true reversal is not expected, but rather a leveling off of observed academic gains concurrent with the removal of the intervention. While this method of analysis allows for the uses of an ABAB with academic intervention, it should be understood that the failure to truly verify initial baseline performance undermines the investigator's ability to rule out alternative conclusions. As such, the ABAB withdrawal design is best suited for behaviors that are sensitive enough to quickly revert to previous levels.

Multiple Baseline Design Several alternative experimental single-case frameworks that allow for verification and replication to be achieved without removal of the intervention have been developed in response to the limitations of the ABAB withdrawal design. As a result, these alternative research designs circumvent the need to consider the ethical, practical, and reversibility challenges previously described. Of these methods, the most widely used is the multiple baseline design, which accomplishes verification and replication through the systematic introduction of the intervention across a number of conditions at different points in time (Cooper et al. 2007). The term *condition* refers to the unit of analysis that will serve as the basis for verifying baseline predictions and replicating intervention effects. That is, multiple baseline frameworks can be conducted across different individuals, behaviors, or settings. Importantly, multiple baselines across behaviors and settings are the only examples that are truly experimental because they meet the condition of the individual serving as their own control. Despite the inability of multiple baselines across participants to provide an experimental analysis of a given intervention, it remains among the most popular single-case designs because it has practical benefits. Regardless of the analysis unit, all multiple baseline designs are essentially a series of AB designs stacked on one another with the introduction of the intervention staggered across each of these defined conditions. Specifically, the multiple baseline design begins with establishing a series of individual baselines with a common start point. Continuous assessment of the target behavior is used to determine if stable response patterns can be established across all baselines. The intervention is then introduced to a single condition with response patterns monitored under both the intervention and ongoing baseline conditions. Decisions to introduce the intervention to a second baseline condition are made only after a clear response has been observed within the intervention condition and stable data in those conditions without intervention.

A critical feature of the multiple baseline design is the time lag between introductions of the intervention to different baseline conditions. These additional observations allow for verification to be achieved through comparisons of individual performance across the various baseline conditions. That is, the response patterns obtained within extended baselines are compared to those in which the intervention was introduced at an earlier point. The prediction associated with the original baseline condition is verified if there was little or no change observed across baseline conditions. Moreover, the time lags also enable replication through comparisons across intervention phases. A reliable change in response patterns associated with introduction of the intervention provides sufficient evidence that the treatment influences the behavior or skill of interest.

Alternating Treatments Design A common limitation of the experimental designs described up to this point is their inability to provide information about the comparability of two procedural modifications. That is, the ABAB withdrawal and multiple baseline designs are most appropriate for evaluating the effects of a single intervention framework. Adapting these designs for investigating issues concerning which intervention will work the best for a particular student or group of students might require a significant time investment. The alternating treatments design provides an experimental approach for comparing behavioral responses to different treatments. Within an alternating treatments framework the procedural modifications being investigated are repeatedly interchanged with the effects on behavior subsequently compared to determine the condition under which responses are most optimal (Kennedy 2005). A notable divergence of the alternating treatments design with the previous experimental frameworks described is that baseline data are not required for an experimental analysis to be conducted. This is because prediction, verification, and replication are not contained within separate phases but rather through the collection of cumulative information obtained through each consecutive data point. Although separate phases are not used to isolate each experimental principle of single-case design, the stability of the obtained data patterns remains crucial for successful evaluation. The data stream obtained for each intervention must have sufficient stability to predict performance in the immediate future, verify previously obtained data, and replicate past performance under the context of the particular procedural modification. Confidence that the more effective intervention is associated with changes in the targeted behavior or skill over and above alternative variables is engendered based on the comparative aspect of the design. That is, clear differentiation between the data patterns obtained from the application of the two treatments provides evidence that improvements are reliably and predictably produced.

Changing Criterion Designs The experimental analysis of interventions designed to increase academic skills or behaviors that may not readily return to baseline levels can be a challenge (Johnston and Pennypacker 2009). Fortunately, the multiple baseline design provides a method for investigating such behaviors. A limitation of the multiple baseline design, however, is the need for multiple units to compare to achieve a truly experimental framework. An alternative method is provided by the changing criterion design, which enables the investigator to examine the effects of some procedural modification on a single behavior or skill. The ability to demonstrate experimental control with a single target behavior is advantageous if a student demonstrates a significant deficit in one specific area. The changing criterion framework requires a graduated intervention approach in which the performance criterion changes systematically. Following the collection of stable baseline data, each new phase represents an incremental adjustment in the performance criterion with successful attainment and maintenance leading to the initiation of a new set of standards. As with all single-case designs, the experimental logic of the changing criterion design is intrinsically related to the stability of the data. Stable data patterns allow for current performance levels to be determined and subsequent predictions to be made. These predictions are verified through examination of the data indicating stable improvements in the target behavior at different criterion levels. Because the design does not allow behaviors to return to original levels, it is often advantageous to enhance verification to improve confidence that the intervention is responsible for performance at different criterion. Increased confidence can be achieved by either varying the number of sessions associated with different criteria or reinstituting previous performance levels. The use of a different number of sessions for each criterion verifies that the target behavior will not increase without further graduation to other criteria. Because an explicit assumption of the changing criterion design is that response levels will not change unless the criterion is altered, the collection of repeated measurements at various criteria provides an opportunity to test this notion with data stability at different levels providing evident confirmation. The verification statement within single-case research can also be enhanced by reinstituting previous criterion levels to demonstrate that responding will return to these earlier levels. Using a direct manipulation of performance criteria provides strong evidence that the behavior or skill would have remained at previous levels if there had been no changes in the criterion.

Visual Analysis The single-case framework discussed in the preceding sections provides a structure for evaluating the effects of an intervention on some targeted behavior or skill. That is, emphasis up to this point has been on the variety of techniques available for arranging conditions to support inferences about whether a particular procedural modification was related to improvements in the outcome variable. The specific designs described allow for several types of evaluation questions to be addressed ranging from basic progress monitoring to more refined assessments of the relation of an intervention with a behavior or skill. It is important to note that the appropriate ordering of conditions and repeated collection of outcome data alone does not allow for successful evaluations. Rather, judgments about whether an intervention is working must be considered within the context of the obtained data.

The traditional method for evaluation in single-case research is accomplished through visual analysis (Horner et al. 2005). Visual analysis refers to a process of data interpretation in which the continual assessment of individual performance on the targeted behavior or skill is graphed and systematically inspected. The goal of visual analysis is to determine whether individual performance on the skill or behavior being targeted differed under varying conditions. This process consists of inspecting both the within- and between-phase data patterns to determine if the intervention is working. The conceptual rationale and graphical features considered for each are briefly described in the following sections.

Within-Phase Data Patterns The importance of the data patterns obtained within each condition for successful evaluation has been addressed throughout this chapter (Gast 2010). That is, much of the discussion on specific single-case designs emphasized the need for stable patterns to assist with making predictions about individual performance in the immediate future. Recall that stable data patterns are those that are characterized by minimal variability and trend that is either flat or moves away from the intended therapeutic effect. The strength of the predictions drawn from the data series is increased with patterns characterized by more consistency. In other words, greater stability within the data increases confidence that future performance will maintain at those levels. This projected performance level serves as the comparison standard for observations taken after the intervention has been introduced. Accordingly, the stability of the within-phase data patterns dictates the quality of comparative evaluations conducted between phases.

Between-Phase Data Patterns The evaluation of intervention effects within a single-case framework requires the comparison of data across different conditions (Gast 2010). These comparisons ultimately provide the basis for determining whether the intervention has been effective or not. Assuming the within-condition data patterns are stable, the trained visual analyst considers a range of graphical features to determine if performance differs across conditions. These graphical features relate to two general characteristics of the data including the magnitude and rate of change across phases. Magnitude is represented by mean and level changes of performance. A mean change refers to shifts in the average performance on the skill or behavior across conditions. Level refers to a discontinuity of performance across adjacent phases. For instance, an abrupt increase in performance that corresponds with an introduction or withdrawal of an intervention indicates that an immediate intervention effect has occurred. The concepts of mean and level change are somewhat related but do not necessarily co-occur. That is, there are circumstances in which a rapid change in level might be observed without an accompanying mean change across phases or that the overall mean changes without a sudden shift in level. Rate of change is represented by changes in trend and the latency of change. Trend refers to a systematic increase or decrease in the behavior or skill over time. The introduction of a new set of conditions might lead to changes in the direction of the behavior possibly indicating consistent improvement or deterioration over time. Even if these effects were not immediate as with a level change, it is likely that such findings indicate important treatment effects. Latency refers to the amount of time that elapses between the introduction of a new phase and changes in performance. The more immediate the performance change occurs following the introduction of an intervention, the more readily the intervention can be tied to those altered patterns. Data evaluations are ultimately based on the magnitude and consistency of these changes across conditions.

Strengths and Limitations of Visual Analysis The appropriateness of visual analysis for evaluation has long been debated in academic circles (Campbell and Herzinger 2010). Most scholars agree that visual analysis continues to be the preeminent method for evaluation of single-case data (Barlow et al. 2009). Within the context of a RTI model, visual analysis is appealing because it represents a widely accessible evaluation method that may be enhanced through further training, but does not necessarily require advanced knowledge of statistical reasoning and interpretation. These strengths are clear when the obtained data patterns provide convincing evidence that the behavior or skill has changed in relation to the intervention. The traditional argument for visual analysis emphasized the need for strong visual effects because it was believed that only those interventions with the most convincing effects would be accepted (Parsonson and Baer 1992). As such, visual analysis was viewed as a means for encouraging both researchers and practitioners to continue to strive towards the development of intervention frameworks that lead to irrefutable treatment gains. Despite these strengths and ambitious aspirations, the utility of visual analysis has continued to be questioned (Maggin and Chafouleas 2013). The primary dilemma for visual analysis is that educational and academic interventions often do not produce data patterns that provide definitive conclusions about treatment effectiveness. Inconclusive data streams accentuate the weaknesses of visual analysis, namely that the process can be highly subjective and lead to inconsistencies between raters. Moreover, the emphasis on interventions that produce only the most definitive effects might lead to the premature discarding of procedural modifications associated with small but consistent results that are difficult to detect visually. These findings might have critical importance for certain populations or settings in which even minor gains are important. These criticisms have led to calls for the development of statistical approaches for describing treatment gains (e.g., Shadish et al. 2013). An advantage of statistical analyses is that subjectivity can be reduced and the consistency of change more objectively evaluated. Unfortunately, the brief, sequential nature of single-case data provides many challenges for obtaining an accurate statistical estimate of treatment gains, though these are beyond the scope of this chapter. Suffice it to say, however, that visual analysis remains the most appropriate method for evaluating single-case data especially within applied settings (Gast 2010). The limitations of visual analysis might be addressed through training of school personnel to systematically inspect sequential data. At the least, it is important to note the potential weaknesses of visual analysis to allow for conscientious use of the technique for evaluation.

Considerations for Successful Adaptation of Single-Case Research Designs The single-case research design provides a useful framework for conducting the evaluations inherent within an RTI model. Specific advantages of the single-

Table 1 Using single-case design and in a response-to-intervention model: some implications for practice

Single-case design is ideally suited for evaluating intervention effects and making data-based decisions in schools. In order to use single-case design in schools, it is key to have a problem-solving team member with expertise in:
The general rules of single-case design
The ability to select an appropriate design
Training and expertise in graphing outcome data using single-case design standards
Training and expertise in visual analysis
Schools should adopt a consistent approach to graphing so that staff members become comfortable and fluent with the data analysis phase
Each intervention case should have each of the following elements:
A defensible measure of the target behavior
A priori selection of an appropriate quasi experimental or experimental single-case design based on the case
An established baseline level of performance
Application of clearly defined intervention
Verification and replication of the intervention effects in cases which require experimental control
A through summary of the outcome data both in terms of the intervention effects and if appropriate the evidence for a demonstration of a functional relationship between the intervention and target behavior

case framework include its history of providing a means for evaluating academic, behavioral, and psychological interventions; its flexibility for addressing a range of evaluation questions that are likely to be confronted in schools and classrooms; and its reliance on continual assessment to provide ongoing formative evaluation of intervention effectiveness. Consequently, the single-case research design is well suited to be adapted for use in schools and classrooms. It is important to note however, that the single-case framework provides only the structure for determining whether a particular practice or strategy is working or not. The utility of using single-case design is fully reliant on RTI teams selecting evidence-based practices in a defensible manner and measuring intervention fidelity (Riley-Tillman et al. 2013). In the absence of well-selected and implemented evidence-based intervention, single-case design is left providing a deeper understanding of the role of unknowable practices. This has limited utility at best. In addition, single-case design is fully reliant on RTI teams employing evidence-based assessment when measuring target behaviors and skills. The outcome data that drives intervention decisions must be both reliable and valid for any decisions made to have rigor. Finally, while single-case design can help identify interventions that are working, it does not identify those that are perceived usable by the teacher/interventionist. It is critical to also assess the long-term viability of an intervention approach so teachers are not overburdened with effective yet debilitating intervention procedures. Finally, in terms of direct use of single-case design in RTI, it is suggested that teams plan for different levels of rigor across tiers. When engaging in whole school, small group, and the initial stages of individual intervention, the primary use of single-case design should be the documentation of a change in target behavior. In these stages of an RTI model, it is not essential to know if the intervention is functionally related to changes in a target behavior. As cases progress and higher stakes come into play, such as eligibility for special education and highly intense intervention practices, fully experimental single-case design should be incorporated (see Table1 Using single-case design and in a response to intervention model: some implications for practice). The use of single-case design at this stage of the RTI process will allow for much more defensible statements about a child response to intervention.

References

Barlow, D. H., Nock, M., & Hersen, M. (2009). *Single case research designs: Strategies for studying behavior change* (3rd ed.). New York: Pearson.

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice*. New York: Guilford Press.

Burns, M. K., Riley-Tillman, T. C., & VanDerHeyden, A. M. (2012). *RTI applications: Academic and behavioral interventions* (Vol. 1). New York: Guilford Press.

Campbell, J. M., & Herzinger, C. V. (2010). Statistics and single-subject research methodology. In D. L. Gast (Ed.), *Single-subject research methodology in behavioral sciences* (pp. 199–233). New York: Routledge.

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J. (2013). Direct behavior rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology, 51,* 367–385. doi:10.1016/j.jsp.2013.04.002.

Christ, T. J., Nelson, P. M., Van Norman, E. R., Chafouleas, S. M., & Chris Riley-Tillman, T. (2013). Direct behavior rating: An evaluation of time-series interpretations as consequential validity. *School Psychology Quarterly*. doi:10.1037/spq0000029.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson Prentice Hall.

Gast, D. L. (2010). Visual analysis of graphic data. In D. L. Gast (Ed.), *Single-subject research methodology in behavioral sciences* (pp. 199–233). New York: Routledge.

Horner, R. H., Carr, E. G., Halle, J., Mcgee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of human behavior research* (3rd ed.). Routledge: New York.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. London: Oxford University Press.

Kennedy, C. H. (2005). *Single-case designs for educational research*. New York: Pearson.

Maggin, D. M., & Chafouleas, S. M. (2013). Introduction to the special series issues and advances of synthesizing single-case research. *Remedial and Special Education, 34,* 3–8.

Parsonson, B. S., & Baer, D. M. (1992). The analysis and presentation of graphic data. In T. Kratochwill (Ed.), *Single subject research* (pp. 101–166). New York: Academic.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York: Guilford press.

Riley-Tillman, T. C., Burns, M. K., & Gibbons, K. (2013). *RTI applications: Assessment, analysis, and decision making* (Vol. 2). New York: Guilford Press.

Shadish, W. R., Kyse, E. N., & Rindskopf, D. M. (2013). Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research. *Psychological Methods, 18,* 385.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Part VII

Contemporary Implementation Topics

Technology-Based Assessment and Problem Analysis

Gerald Tindal and Julie Alonzo

This chapter discusses the features of a widely adopted technology-based assessment commonly used in schools implementing response to intervention (RTI). First, the chapter describes easyCBM® in the context of RTI. An important perspective is that a number of dimensions need to be addressed such as multiple references for decision-making, flexibility in measurement implementation, and documentation of technical adequacy. The second section addresses specific research conducted in the development of easyCBM® for reading and mathematics (organizing the discussion by grade bands: kindergarten to grade 2 and grade 3 through 8). Finally, this section ends by noting the strengths and utility of technology-based assessment to enhance decision-making. A number of reports are discussed that follow a sequence of teacher decision-making: (a) determining risk and grouping students, (b) diagnosing instructional needs of students and devising instructional plans, (c) evaluating these interventions using time-series progress monitoring, and finally (d) evaluating programs for both individuals and systems using grade-level and building-level changes, though a number of other variables could be considered (e.g., race–ethnicity, English language learners). The third and last sections of the chapter both reflect on critical dimensions of technology-based assessment and speculate on future directions in this field. After articulating a conceptual model for easyCBM® using three interlocking critical features (measurement sufficiency, instructional adequacy, and data-driven decision-making), we venture into the need for professional development with effective reports to catalyze training enhancements.

G. Tindal (✉) · J. Alonzo
University of Oregon, Eugene, USA
e-mail: geraldt@uoregon.edu

Overview and Description of easyCBM®

Technology-based assessments have increasingly become essential tools for schools and districts implementing RTI. The best examples of technology-based assessment systems for use in RTI approaches share many features.

1. The process is efficient for staff to roster their students and activate the program with a minimal amount of work and yet allow flexible grouping of students, with the ability for teachers to regroup students throughout the year as individual needs change in relation to their peers.
2. Online platforms provide easy access to test themselves as well as the possibility of embedded trainings on test administration.
3. Well-designed databases facilitate immediate reporting of results, with historical records allowing districts to track the effect of different approaches over time. Such systems capture detailed information about the interventions programs and outcomes that provide critical information to guide lesson planning.

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_28

4. The system is easy to use as the more complex computer programming is in the background and entirely managed using well-designed computer interface for all student and teacher interactions.
5. The system is designed for incremental enhancements to take advantage of advances in both measurement and technology.

This chapter focuses primarily on easyCBM®, one such technology-based assessment system, designed as part of a federally funded National Center on Progress Monitoring. Although other technology-enhanced assessment systems share some of the same features, easyCBM® is used as the illustrative example because we are most familiar with its features and the principles that guided its design and fuel its continued enhancements.

Initially introduced in the fall of 2006, easyCBM® has continued to grow in popularity and use. At the time of this writing, over 325,000 educators, representing over 2.3 million students, have established accounts with the online learning system, with accounts present from every US state as well as a number of international locations. Over 17.4 million easyCBM® tests have been taken since the system was first made available. Many school districts have incorporated easyCBM® as an integral part of their RTI protocols. Student performances on the easyCBM® benchmark measures are used to identify students for additional intervention, and their scores on the easyCBM® progress-monitoring measures are used, in part, to evaluate the effectiveness of provided interventions and to modify instruction as needed. For academically struggling students, lack of progress on curriculum-based measurement (CBM) measures, when the students have been provided with appropriate intervention, implemented with integrity for a sufficient period of time, has served as the primary factor for determining eligibility to receive special education services.

CBMs Applied in the Context of RTI

CBM assessments include both individually administered measures (e.g., phoneme segmenting, letter names, letter sounds, word reading fluency, and passage reading fluency) and group-administered measures (e.g., vocabulary, comprehension, mathematics). Test administration procedures are typically described in a teachers' manual. On the easyCBM® system, the utility of the teachers' manual, available as a downloadable portable document format (PDF) directly from the easyCBM® site, is bolstered by the addition of an online training link, with videos on standardized test administration and scoring and proficiency checks of mastery to be completed before a person administers the tests. For greatest utility, assessment systems designed for use in RTI contexts must include both screening and progress-monitoring measures.

As is typical of such assessment systems, the easyCBM® system includes both universal screening measures (for fall, winter, and spring administrations) and multiple alternate forms used in evaluating instruction and progress. Specific measures included on the screening assessments vary by grade level, with the measure type selected based on empirical findings of which tests provide the most robust screening of content and skill at each grade level.

An array of progress-monitoring assessments should be available at each grade. These measures provide in-depth information about individual students' particular strengths, weaknesses, and needs for tier 2 supplemental instruction and tier 3 intensive intervention. Progress-monitoring forms typically include not only the specific measure types included on the universal screening assessment but also go beyond these measures to provide information about a wider swath of skills. These diagnostic measures should not only provide relevant information about specific skill areas with which students are struggling but they should also be optimized to be sensitive to growth, enabling teachers to evaluate the impact of their instruction and modify it when warranted. In the most useful technology-based systems, intervention data are logged directly into the system, which plots intervention lines on the individual student graphs, providing individualized histories of student response to instruction or tiered supports.

When systems truly capitalize on the strengths of technology-based assessments, the group-administered measures from both universal screening and progress-monitoring assessments can be administered via desktop computers, laptop computers, or tablet devices such as iPads. Individually administered measures can be administered using paper and pencil, with scores later entered online, or they can be administered directly from tablet devices, to streamline data collection.

Technical Adequacy Considerations

In keeping with the original precepts of CBM, the measures need to be sensitive to the instructional needs of students with disabilities as well as to students from the general education population. Because of the need for alternate forms, measures need to be appropriately scaled, a consideration that can be addressed during measurement development with the use of item response theory (IRT). Finally, traditional reliability and validity requirements need to be met, as promulgated by the 2014 *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014).

For example, all easyCBM® measures were developed using the principles of universal design for assessment (Thompson et al. 2005) and followed the guidelines for test development as described in *The Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 2014) with particular attention to accessibility and freedom from bias. During item development, teams of experienced grade-level educators were hired to draft the measures, with thorough review by a team of trained assessment researchers at the University of Oregon. Following content and bias/sensitivity review, items were piloted with grade-level students (with exact sample size varying by measure and grade). IRT modeling was used to analyze pilot data, with item information (measure, standard error, discrimination, mean square outfit) used to create multiple alternate forms for screening and progress-monitoring measures.

Screening assessments are typically optimized for the purpose of identifying students who need additional supports to meet grade-level content and performance standards in the areas of reading and mathematics. Progress-monitoring assessments, on the other hand, may be designed to provide more detailed diagnostic information related to specific skill deficits and developed to ensure comparability of alternate forms and sensitivity to growth for students receiving targeted instruction in specific skill areas.

Following initial form creation, additional studies are typically conducted to evaluate and document the reliability (test–retest, alternate form), internal consistency (both within and between measure types), sensitivity and specificity, generalizability of the measures, and to provide evidence of their validity in making screening- and progress-monitoring decisions. Results of these studies should be detailed in technical reports or other published materials readily available from the CBM developers. For instance, over 100 technical reports documenting easyCBM® measurement development and technical adequacy studies are available on the website of behavioral research and teaching (BRT) at the University of Oregon (brtprojects.org). The need to consider the technical characteristics of the measures is emphasized because it is believed that instructional planning and evaluation is not a low-stakes decision that can be made on the basis of informal measures. Rather, it is a high-stakes decision that needs to use measures that have all the characteristics demanded by the test standards.

Technology-Enhanced Reading Measures In kindergarten, first, and second grade, early literacy measures focus on letter names, letter sounds, phoneme segmenting, and word-reading fluency (Alonzo and Tindal 2007); in grades 2 to 8, vocabulary and direct measures of comprehension measures are available (Alonzo et al. 2012a-g; Irvin et al. 2011, 2012a, b; Lai et al. 2010; 2012a, b; Park et al. 2012a, b). Although other approaches are possible, use of IRT enables creation of item pool with known item difficulty, facilitating the creation of comparable alternate forms. For instance, use of IRT during measurement development resulted in the easyCBM® early literacy measures alternate form reliability ranging from 0.86–0.91 for the Phoneme Segmenting measures, 0.82–0.89 for the Letter Names measures, 0.76–0.88 for the Letter Sounds measures, and 0.95–0.96 for the Word-Reading fluency measures in grades K-2.

In the late elementary and middle school grades, a variety of reading measures have been developed and investigated, including word and passage reading fluency, vocabulary, and comprehension. Some CBM systems use maze tasks for measuring comprehension, while others use passages followed by selected response items. As with the early literacy measures described earlier, use of IRT during measurement development provides insights into item characteristics instrumental in creating assessments that are sensitive for use as screeners as well as for monitoring the progress made over time. With a view toward ensuring the utility of the assessment for students at risk, developers may focus on building a test that begins with a sufficient number of easily accessible items to establish an accurate base of knowledge or skill, yet also includes items that cross the range of difficulty, providing the means by which to sensitively measure improvements over time for students with a range of skill/knowledge. Measures that contain an insufficient number of "easy" items, or alternatively that contain an insufficient number of "moderately difficult" and "difficult" items may not be useful for screening purposes, and also fail to serve the dual purpose of providing information about progress over time.

Technology-Enhanced Mathematics Measures Although early work in mathematics CBM focused on fluency-based short probes, more recent developments reflect the increasing depth of knowledge and skill expected of students, with a greater emphasis on items aligned to state and content standards, such as the Common Core State Standards (CCSS) in Mathematics and the National Council of Teachers of Mathematics' (NCTM) Focal Point Standards.

The easyCBM® mathematics progress-monitoring measures were initially developed to be sensitive for students with persistent learning problems (Alonzo et al. 2006; Anderson et al. 2010, 2011), using the NCTM Focal Point Standards to establish the content and performance expectations used (Alonzo et al. 2010; Nese et al. 2010a, b). At the time this chapter was written, the easyCBM® mathematics measures were being revised with (a) alignment to the CCSS for the existing measures and forms and (b) development of new items and forms written in alignment with the CCSS. As in the reading measures, the same Rasch scaling was used to create equivalent alternate forms in grades K-z (Alonzo et al. 2009b, c, d). A series of technical reports were written describing the process of developing items, articulating the blueprint and scaling process as well as the development of forms: grade 3 (Alonzo et al. 2009e), grade 4 (Alonzo et al. 2009f), grade 5 (Lai et al. 2009c), grade 6 (Lai et al. 2009b), grade 7 (Lai et al. 2009b), and grade 8 (Lai et al. 2009d). Both students and items were placed on the same scale using a 1-parameter Rasch model; the results showed the average for items was lower than the average for students and alternate test forms within each grade had a mean difficulty (IRT *Measure*) within 0.20 Rasch units of one another (Anderson et al. 2011). Internal consistency ranging from 0.78–0.91 was documented once the alternate forms were completed (Anderson et al. 2009, 2012a, b). Validity also has been established (Anderson et al. 2011a, b, c, d).

Additional Technical Adequacy Considerations

School districts increasingly use technology-enhanced assessments to identify students who are at risk of not passing their statewide assessments. For this application, studies establishing the utility of the assessments for that purpose are crucial. Normative performance benchmarks can provide insights into how a student's performance compares to grade-level peers (Tindal et al. 2009a, b). A variety of criterion-related validity studies documenting the sensitivity and specificity of CBM assessments for predicting performance on statewide large-scale assessments have been published. Such studies typically include correlational as well as multiple regression analyses (Anderson et al. 2010a, b). In addition, it is becoming more common for assessment developers to report on the diagnostic efficiency of such measures. For example, the diagnostic efficiency of the easyCBM® mathematics measures was determined using a receiver operating characteristic (ROC) analysis; ROC analyses were cross validated with unique samples. In a ROC analysis, classification of students (using state test proficiency status) is used to determine sensitivity (true positive proportion) and specificity (true negative proportion) of the screening (benchmark) measures. A ROC curve plots the screening test's false-positive rate on the horizontal axis and sensitivity on the vertical axis for $k-1$ scores where $k=$the total number of unique scores occurring on the screening measure. Thus, each point on the curve depicts the sensitivity and false-positive rate for every score in the range of scores on the test measure (http://www.ajronline.org/content/ 184/2/364.full) reflecting the reality that gains in sensitivity always come with increased false-positive error rates. Area under the curve (AUC) curve are used as a measure of predictive power. Results indicate that the easyCBM® math measures function well, with AUC results ranging from 0.85 to 1.00 across grades K-8 (Anderson et al. 2011a, b).

In summary, technology-enhanced assessment systems such as easyCBM® have benefited from years of research funded by the Institute of Education Sciences (IES) and, earlier, the Office of Special Education Programs. Technical reports are available on the BRT website: http://brtprojects.org. This ongoing and systematic program of research continues to inform refinement of such assessment systems. In-depth longitudinal study of the measures in use, along with regular feedback solicited from practitioners using the measures in their local RTI contexts ensures continued development and adaptation to meet the needs of educators.

Strengths of Technology-Based Assessments

Technology-based assessments offer many benefits to users including increasing the security and efficiency of data collection efforts, streamlined processes for sharing results, and savings of time and money by reducing the need for printing and shipping testing materials. In addition to these more obvious benefits, technology-based RTI systems can offer some more subtle insights, such as facilitating deeper understanding of student performance data, fostering a team approach to meeting student needs, and harnessing the power of databases to assist in identifying patterns in the data that might not be so readily observable from visual inspection of raw scores.

Similar to other technology-based assessment systems, the easyCBM® system uses secure uniform resource locators (URLs), with administrator control of access to data. District-level users are able to access performance information from all users in a school district; building-level users see performance data from all students assigned to their school, and classroom-level users can access data only for students specifically associated with them in the database. Fluid-grouping features enable multiple users to access the data and form groups for analysis/interpretation without disturbing other users' organizational structure. These features enable a Title 1 coordinator or a special education professional, for example, to access student performance records for students in their caseloads, even when others (e.g., classroom teachers, instructional assistants) administer the tests. Each user is assigned a unique user name and password (which they can update to maintain security), and data are encrypted such that they cannot be viewed except when using the reports on easyCBM® or exported in comma separated values (csv) format in preparation for upload to a district or state student information system.

Reports on the technology-enhanced assessment systems are designed to provide useful information to guide decision-making to meet different users' needs. At the district level, reports can facilitate discussions related to resource allocation and program evaluation, enabling district administrators to document the impact on student learning of district initiatives and identify areas of greater/lesser need, and specific skill areas for which professional development or additional supports might be needed to enable students to meet rigorous content and performance standards. At the school level, reports enable users to identify specific broad constructs (vocabulary, fluency, comprehension, mathematics) in which students are struggling, thus facilitating decisions related to programmatic and curricular supports. Classroom-level reports provide insights into the specific skills students have mastered or with which they are struggling (e.g., specific letter sounds, particular objectives within math, or inferential rather than literal comprehension), fostering informed lesson planning based on student needs. Individual reports enable teaching teams to monitor the effectiveness of specific interventions for individual students and groups of students and provide an accessible and efficient way to communicate with parents.

One of the greatest benefits of using a technology-based assessment system is the increased functionality it provides in terms of tracking student responses and providing useful feedback to teachers and students themselves. Technology-enhanced assessment systems optimize group-administered measures for online testing, with student responses captured in real time as students complete the tests, enabling instant score reporting upon completion of the assessment. Carefully thought-out systems can enable teachers to track student completion of assessment items as they are working their way through the tests, and can provide a variety of reports and graphs to help teachers interpret student performance without having to engage in additional data entry or recording of student data. Programs can include detailed reports for students, helping them track their own progress in specific skill areas and providing instant feedback on their performance, enabling them to identify areas of confusion that would benefit from additional attention in their studies.

Well-designed computer-administered assessments can also streamline the testing experience for students, with accommodations built into the infrastructure such that they can be delivered with a minimum of demand on school resources. For instance, on assessments that do not target reading, audio files can be embedded within the website, enabling students to access a read-aloud accommodation without needing to be singled out for individual attention or one-on-one administration. All of the easyCBM® mathematics items that include words, for instance, include a read-aloud option. And, as of January 2014, all easyCBM® mathematics items include an optional Spanish language accommodation enabling district administrators to provide Spanish language accommodations (both written text and read aloud) for every mathematics item K-8. Again, forethought about programming is essential. In the case of the Spanish language accommodation, for instance, it is important that administrators be able to deactivate this option in states where Spanish language accommodations are not allowed, and reports must be programmed to record whether the test was taken in an English-only or a language-accommodated condition.

In addition, with intentional planning and design, features such as careful use of blank space to enhance accessibility and attention to the fine motor skills required of students as they are interacting with an online assessment can be integrated throughout. Such design considerations play an important role in reducing construct-irrelevant variance that might be introduced for students who lack dexterity with the computer or who have visual challenges, reducing the likelihood that such student characteristics will cause scores to be artificially deflated. The technology used in some online assessment systems enables the measures to be made available as downloadable PDFs. The ability to download the measures for students who have difficulty accessing them in their online format facilitates local accommodations to meet IEP requirements and can also accommodate settings with limited computer access. Many online tests can be taken in a single sitting or administered in shorter segments to accommodate student needs. When this technology is included, student responses can be retained in the system, with the computer automatically returning the student to the test at the point at which he/she stopped, if multiple testing sessions are used.

Ways Technology-Based Assessment Enhances Decision-Making

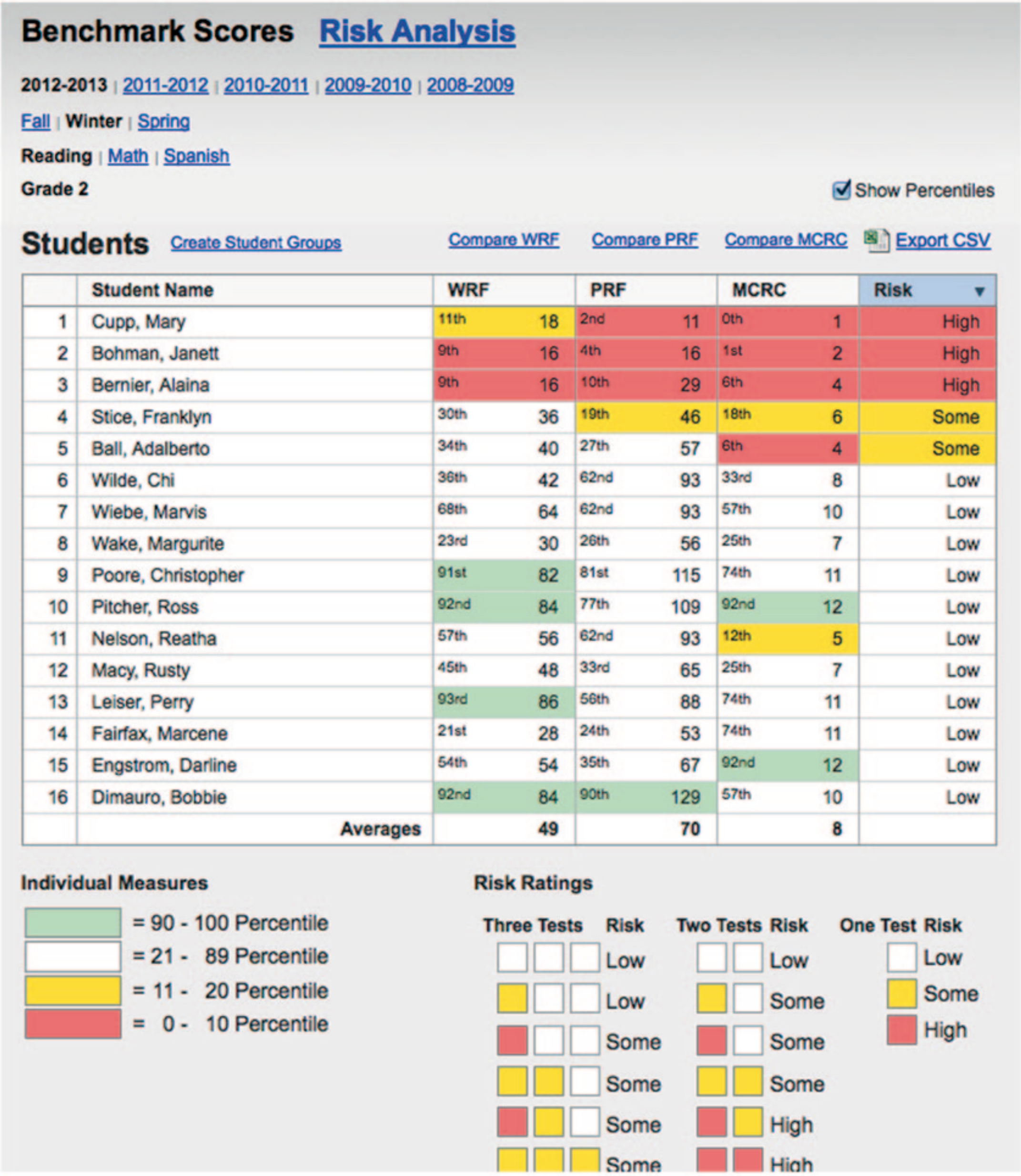
Computer-administered assessments offer additional benefits in the form of instantaneous scoring and report generation. When attention is paid to the need for quick access to testing results, systems can be programmed to provide reports as soon as students finish their assessments. Computer-administered assessments can also streamline the process of providing opportunities for students to make up tests they have missed due to absence, enabling teachers to administer the make-up assessments in a group or individual setting, based on which provides the most benefit in a given situation.

Thus, well-designed technology-based assessments can provide criterion-referenced information (as with the item-level analysis shown in Fig. 2), individually referenced information (as in the individual progress monitoring graph shown in Fig. 4), as well as norm-referenced information (as shown in Figs. 6–8). Benchmark screening reports, such as the one displayed in Fig. 6, help teachers quickly identify which skills individual students are struggling with as well as to identify patterns in their classes as a whole. Usability features such as making the data sortable from high to low or low to high by clicking on the column headers enhance the utility of such reports. The next section describes a sequence of decisions that technology-enhanced assessments facilitate.

Risk Analysis and Group Assignment One of the initial decisions that need to be made at the beginning of the school year is to determine who is at risk for learning problems. For this decision, benchmark measures are administered to all students during fixed periods of time (fall, winter, spring). These common assessments allow teachers to compare students to each other and determine where they fall in a performance distribution. Students who are performing below district-set percentile rank (PR) cut points can be identified. When a comparatively low level of performance appears across different measures, it likely represents an overall need for extra support, classified on the easyCBM® system as a *risk rating*. In Fig. 1, students are grouped on the easyCBM® fall measures into four groups: (a) below the 20th PR in all three areas measured, displayed as red; (b) below the 20th PR in a couple of measured areas, displayed as yellow; and (c) not below the 20th PR in any area measured, displayed with no color; and (d) exceptionally high performance (e.g., 90th PR) in any measured area, displayed in green. In the end, a risk classification is given as *high* (those marked red), *some* (inconsistently red and yellow), or *low* (few to no areas of risk). See Fig. 1 for a sample display of students in a teacher's classroom.

Group reports help teachers make decisions about how to organize their students for instruction. Figure 2, for instance, displays the performance of a whole class on the fall reading comprehension test. The bar graph highlights the difference in proficiency in reading comprehension of students in the particular class, enabling teachers to identify the three students who may benefit from a more intensive intervention (or may need a fluency-building intervention to assist in comprehension) as well as the seven students whose performance indicates a need for targeted comprehension instruction.

Design of Instruction with Attention to Detail Widespread access to educational technology has expanded the universe of possibilities for effective assessment systems in support of RTI. Some features of technology, in particular, are worth highlighting as they relate to assessment considerations. Computers and tablet devices such as iPads provide opportunities to capture item-level data that can provide additional instructionally relevant information, over and above raw or scaled scores. For instance, it is possible to use item-level responses, captured



Benchmark Scores Risk Analysis

2012-2013 | 2011-2012 | 2010-2011 | 2009-2010 | 2008-2009

Fall | **Winter** | Spring

Reading | Math | Spanish

Grade 2 ☑ Show Percentiles

Students Create Student Groups Compare WRF Compare PRF Compare MCRC Export CSV

	Student Name	WRF		PRF		MCRC		Risk ▼
1	Cupp, Mary	11th	18	2nd	11	0th	1	High
2	Bohman, Janett	9th	16	4th	16	1st	2	High
3	Bernier, Alaina	9th	16	10th	29	6th	4	High
4	Stice, Franklyn	30th	36	19th	46	18th	6	Some
5	Ball, Adalberto	34th	40	27th	57	6th	4	Some
6	Wilde, Chi	36th	42	62nd	93	33rd	8	Low
7	Wiebe, Marvis	68th	64	62nd	93	57th	10	Low
8	Wake, Margurite	23rd	30	26th	56	25th	7	Low
9	Poore, Christopher	91st	82	81st	115	74th	11	Low
10	Pitcher, Ross	92nd	84	77th	109	92nd	12	Low
11	Nelson, Reatha	57th	56	62nd	93	12th	5	Low
12	Macy, Rusty	45th	48	33rd	65	25th	7	Low
13	Leiser, Perry	93rd	86	56th	88	74th	11	Low
14	Fairfax, Marcene	21st	28	24th	53	74th	11	Low
15	Engstrom, Darline	54th	54	35th	67	92nd	12	Low
16	Dimauro, Bobbie	92nd	84	90th	129	57th	10	Low
	Averages		49		70		8	

Fig. 1 Group benchmark report

via online test administration, to provide reports that document specific areas of weakness or strength for students in a given classroom. Figure 3, for example, shows a report from one of the easyCBM® comprehension measures. This report gives the teacher insight into the need to focus more on inferential and evaluative comprehension than on literal comprehension when planning lessons, and also highlights the specific error patterns shown by different students in the class. Computer programming within the system enables a teacher to identify patterns in individual student responses: When a teacher "hovers" over a student's name in the list, the name is highlighted, making it easier for teachers to find

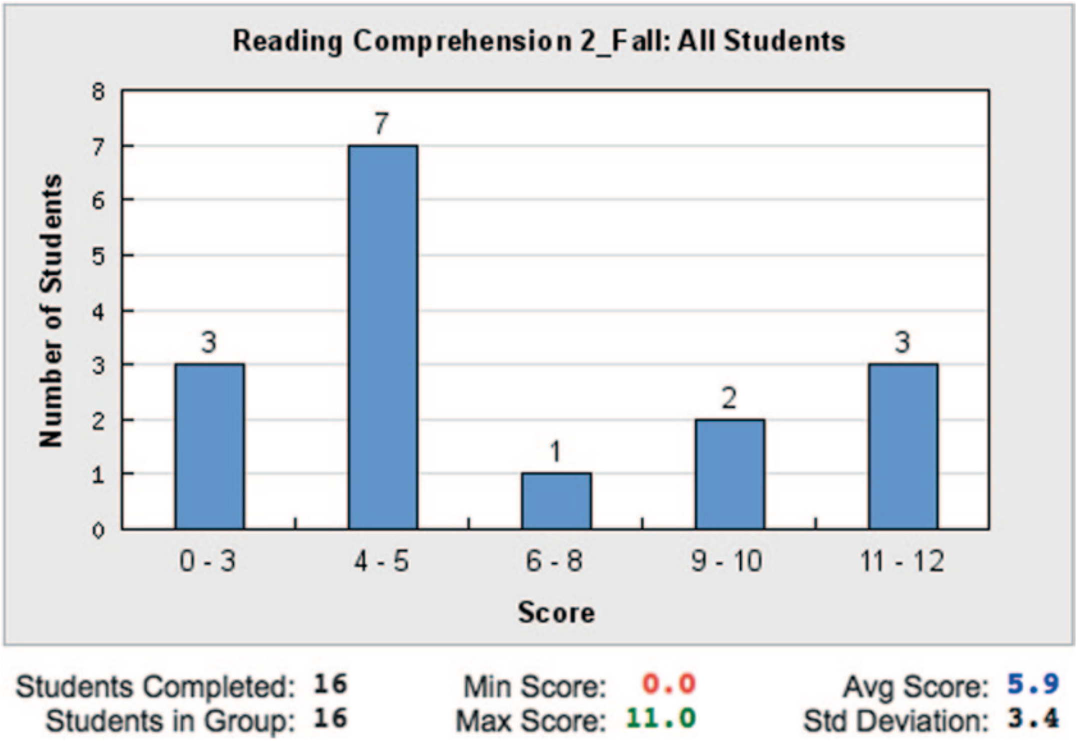


Fig. 2 Group report on a fall benchmark screener assessment

patterns in student responses they might not otherwise notice.

Intervention Evaluation A key feature of RTI is the focus on individual students' response to interventions. Technology facilitates the gathering and interpretation of such information. Figure 4 shows a screen shot from an easyCBM® individual student progress report.

Student scores are plotted over time, enabling the teacher to evaluate growth across the school year. Colored lines on the graph provide normative referents, enabling quick evaluation of the degree to which the student's performance is at grade level (meets the 50th percentile line plotted on the graph) or falls below expectations (to the degree that a student's performance falls further below the 50th percentile, his/her level of need can be interpreted to be greater). These individual reports also provide a place to record the interventions a student has received, with the database providing a convenient and cost-effective historical archive of these data, so subsequent years' teachers have ready access to the instructional approaches that have proven effective—or ineffective—for a given student. The utility of this historical record that combines details about the interventions provided and a record of student performance over time cannot be overemphasized. Without the log of interventions (See Fig. 5), assessment data are at best, difficult to interpret and at worst, relatively meaningless.

Program Evaluation for Individuals and Systems Figure 6 documents an individual student's assessment performance history against the backdrop of the district's performance on the same assessments. At a glance, one can see that the example student Adalberto's oral reading fluency in the fall of second grade (as measured on both the Word and Passage Reading Fluency measures) was near grade-level expectations, with scores that placed him at the 45th and 59th percentile, respectively, when compared to national norms, while his performance on the comprehension measure in the fall placed him near the cutoff

Item Analysis

Easiest to Hardest Items				
Item	Type	Students Correct	Percentage	Student Names, Incorrect
6	Inferential	4 of 4	100%	
5	Literal	4 of 4	100%	
1	Literal	4 of 4	100%	
12	Inferential	3 of 4	75%	Christia T
19	Literal	3 of 4	75%	Denver S
14	Literal	3 of 4	75%	Denyse F
16	Evaluative	3 of 4	75%	Christia T
15	Inferential	3 of 4	75%	Denyse F
8	Literal	3 of 4	75%	Denyse F
3	Inferential	3 of 4	75%	Christia T
2	Literal	3 of 4	75%	Christia T
17	Inferential	3 of 4	75%	Christia T
4	Inferential	2 of 4	50%	Gail P, Christia T
7	Evaluative	2 of 4	50%	Denyse F, Christia T
20	Evaluative	2 of 4	50%	Denyse F, Denver S
18	Evaluative	1 of 4	25%	Gail P, Denver S, Christia T
10	Evaluative	1 of 4	25%	Gail P, Denver S, Christia T
9	Inferential	1 of 4	25%	Denyse F, Gail P, Christia T
11	Literal	0 of 4	0%	Denyse F, Gail P, Denver S, Christia T
13	Evaluative	0 of 4	0%	Denyse F, Gail P, Denver S, Christia T

Fig. 3 Item analysis report to target instruction

point for being identified as at risk in his district, with a 23rd PR. By the winter benchmark assessment, although Adalberto's raw scores continued to improve, his standing in relation to his peers dropped, placing him between the 27th and 34th percentile in oral reading fluency and at the 6th percentile in reading comprehension. This report can help prompt discussion and problem-solving on the part of Adalberto's teachers, as they work together to put together an instructional plan that will help him reverse this downward trend in performance.

Technology-based assessments constructed on a foundation of well-designed databases facilitate district-level decision-making as well. Figure 7 depicts a school comparison evaluation report that provides administrators with a wealth of information in a single screen. At a glance, they can identify the schools in their district with the most intense need for academic support structures, and by toggling back and forth between the "Counts" and the "Percentages" options, they can quickly gather data that might be relevant to share with stakeholder groups, including school boards and, potentially, outside agencies that might offer funding opportunities to support innovations. The links on the page enable users to focus on specific grades or all grades concurrently for content areas. Users can click on school name to dig deeper into patterns they want to explore based on the initial review of the data. Within the Building Level report, district administrators can review the progress students are making over the year, by class.

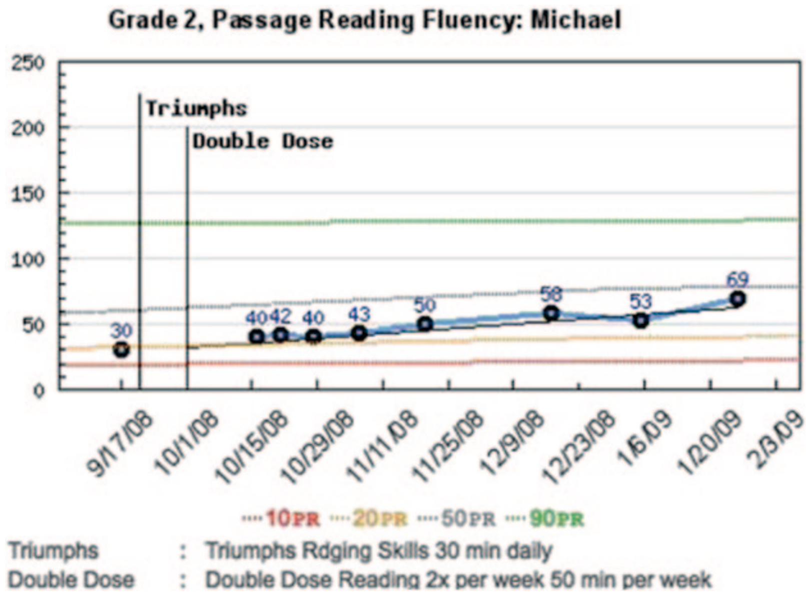


Fig. 4 Individual student progress report to evaluate instruction

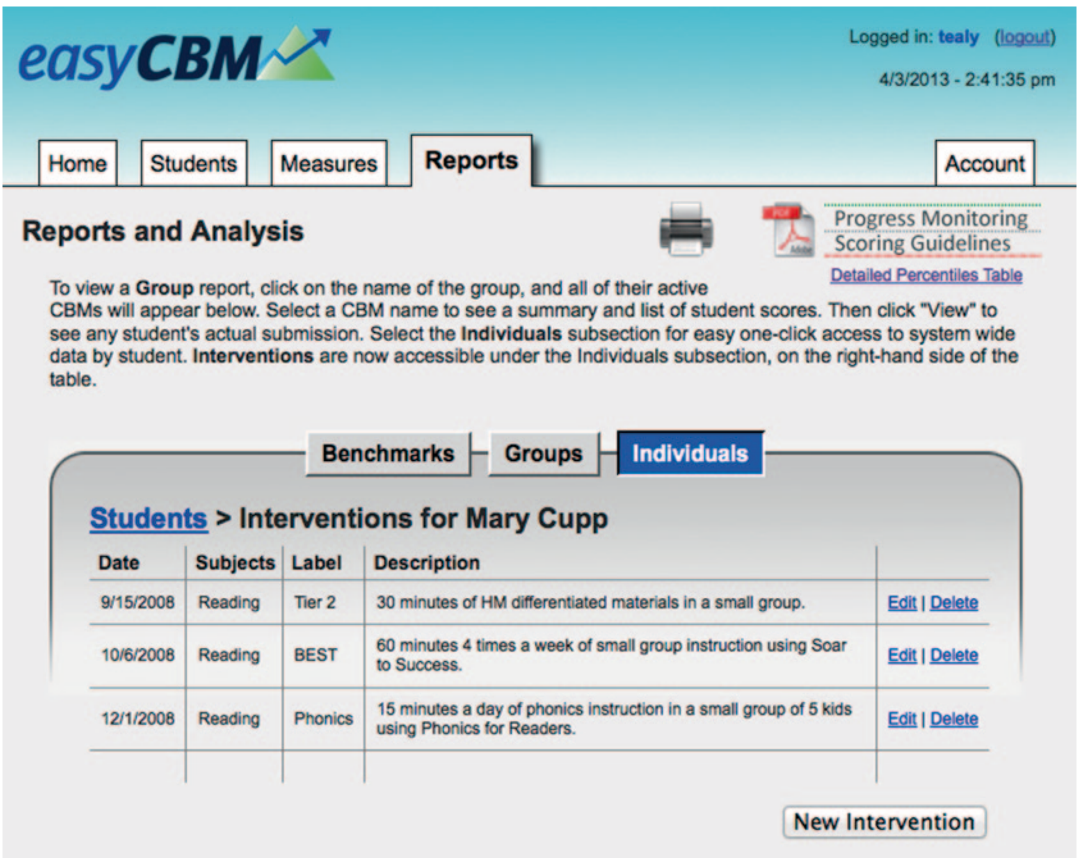


Fig. 5 Detailed log to label and describe instruction for each student

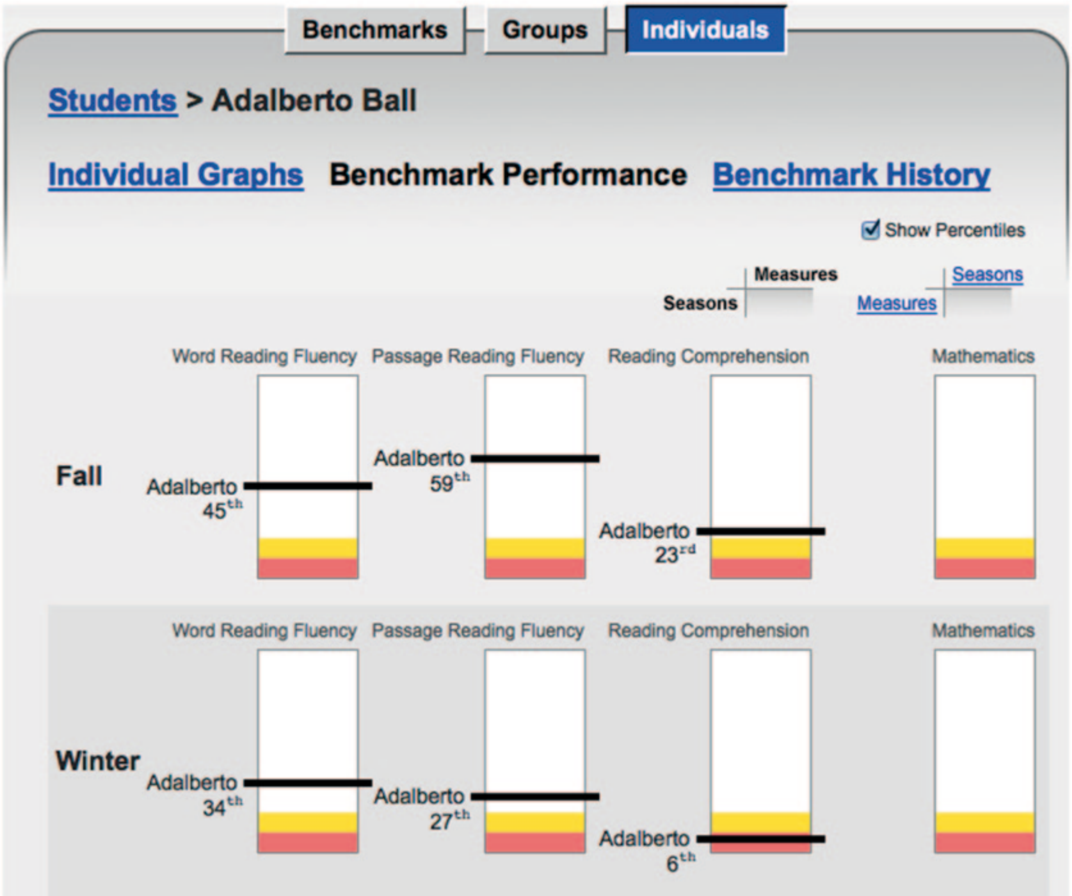


Fig. 6 Individual benchmark history report

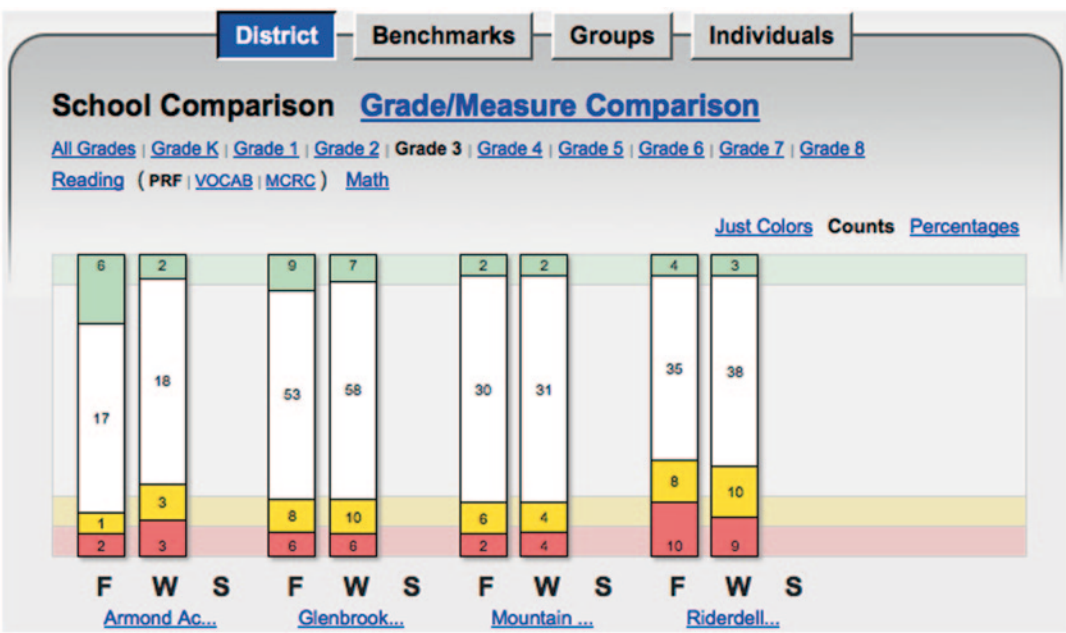


Fig. 7 Systems evaluation report with school outcomes

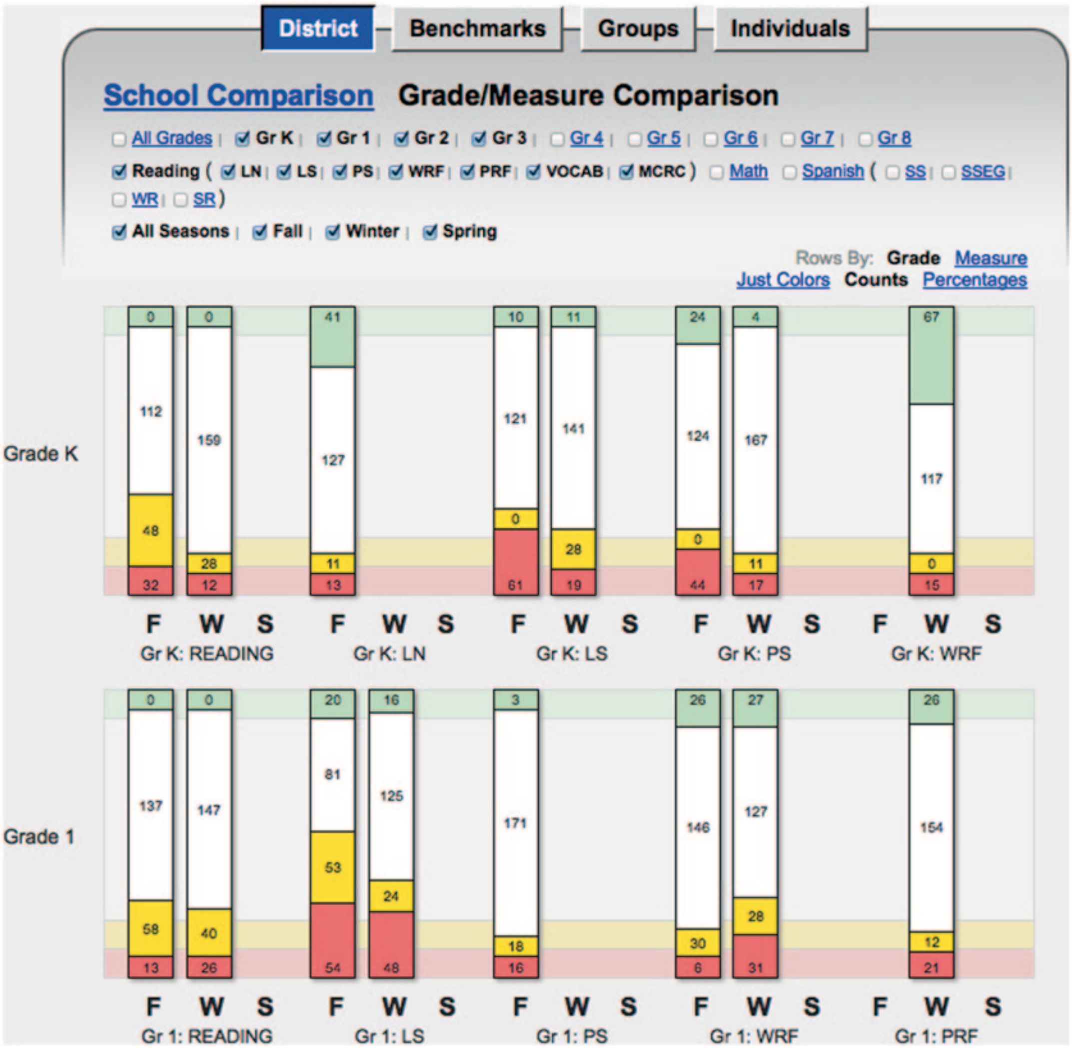


Fig. 8 Systems evaluation with grade/measure outcomes

Figure 8 displays the grade/measure comparison report available for district-level users. Using this report, administrators can look for patterns in their overall district performance, organized by grade level and season, with the ability to select measures. Color-coding, based on the district's preset risk ratings, facilitates interpretation of the reports.

In summary, well-designed technology-based assessments can provide norm-referenced information like the benchmark screening reports, as was displayed in Figs. 1 and 2, helping teachers quickly identify who is at risk, facilitating grouping of students to maximize instructional opportunities. Once these macro outcomes are accomplished, further reports can be generated to target specific skills with which individual students are struggling as well as to identify patterns in their class as a whole, providing criterion-referenced information (shown in Fig. 3) that can be used to document specific instructional programs (Fig. 5). At this point, formative assessment can be used to evaluate instructional programs using individually referenced information (as in the individual progress monitoring graph shown in Fig. 4). After a period of time (perhaps within each seasonal assessment), programs can be evaluated for individuals using norm-referenced information (as shown in Figs. 6–8).

Reflections and New Directions

The essential features of progress-monitoring assessments, often referred to as CBMs, have changed very little since the 1970s. Such measures are still expected to sample from a year's worth of curriculum. They are meant to provide teachers with meaningful information about the progress students are making in mastering that material. In addition, to enhance their utility, progress-monitoring measures are intended to be easy to administer, score, and interpret. See Tindal (2013) for a historical summary of CBM.

However, whereas four decades ago, researchers deemed CBMs as not requiring any particular expertise to develop, the increasing stakes associated with assessment results as well as advances in psychometrics have significantly altered this perspective: It is now recognized that the creation of reliable and valid progress-monitoring measures requires specialized knowledge beyond what most public school teachers possess. This realization spurred the creation of "next-generation" CBMs, measures created using rigorous alignment with standards and stringent statistical modeling (Alonzo et al. 2006). In all of our measurement development, the authors have worked with expert teachers throughout the country to develop reading and mathematics benchmark and progress measures.

Probably the most important advancement with easyCBM® and other modern technology-enhanced assessments is the use of IRT (Embretson and Reise 2000) during test construction, which sets these measures apart from more traditionally designed formative assessments (aimsweb and Dynamic Indicators of Basic Early Literacy Skills; DIBELS). With IRT, test developers are able to calculate measure difficulty for every item and then develop forms that are equivalent. In using IRT, test developers also have been able to place items and students on the same scale to also ensure sensitivity to progress. Such measures stand out for their sensitivity in monitoring growth, the stability of alternate forms, and the provision of measures suitable to assess the full range of student skill in critical content areas from kindergarten through eighth grade.

Another significant difference between easyCBM® and most other progress measures has been the focus on alignment with standards. By design, items have been built in both reading and mathematics that are aligned with standards. When easyCBM® was initially developed, the National Reading Panel's report was used to develop measures in reading and the NCTM Focal Point Standards for measures in mathematics. More recently, however, new items have been developed that are aligned with the CCSS in both content areas. Formal studies of the alignment between these national standards and the assessments document the measures' appropriateness for use in standards-based school systems.

Conceptual Model for Impacting Learning with Technology-Enhanced Assessments Although scaling and alignment form two recent innovations into progress monitoring with easyCBM®, the most significant innovation is the movement beyond a measurement model to an instructional model for which progress monitoring is meant to be used in evaluation (Tindal 2013). The conceptual model for easyCBM® is based on three warrants designed to optimize RTI:

1. Assumption 1 (measurement sufficiency): Students are appropriately placed in long-range goal material to ensure the measures are sensitive to change: What is the type of measure, grade level of measure, and the time interval (density of measures) used during the year?
2. Assumption 2 (instructional adequacy): Instruction is detailed and explicit, allowing a team of teachers to coordinate various elements such as providing an instructional tier (1–3), allocating sufficient time to teach, grouping students appropriately, deciding on the instructional emphasis (alphabetic principles and phonemic awareness, decoding, fluency, vocabulary, and comprehension), using specific curriculum (core and supplemental)

materials, and determining what instructional strategies to use.

3. Assumption 3 (decision-making): Interventions need to be introduced for low-performing students when data warrant change. Are interventions provided at the right time and in accord with specific data features (e.g., level, variability, and slope)?

The theory of change in using technology-enhanced assessments to drive improvement in learning is best reflected as the interlocking union of the three components in a chain: measurement sufficiency, instructional adequacy, and decision-making. It is not each link itself that is critical, but the intersection of the link with subsequent links. As teachers collect data (from benchmark to decisions of risk and monitoring of progress), the data used to inform them need to be sufficient, directed toward instruction, and adjusted as needed (flexible and prescriptive). Furthermore, this information needs to be collected into a single database for teachers to monitor their application of RTI as well as policy-makers and administrators to use the information in making systems-level decisions.

The conceptual model is driven by accessibility as a key ingredient to change: If information is not easily accessible and tractable, then it is unlikely to result in use. The theory is also driven by a holistic approach to change: Changing individual components as separate events is unlikely to change systems. Rather, the whole needs to be reflected in the parts that in turn need to connect teachers and administrators. Finally, research in practice is needed. By using a developmental process, the influence of all components can be modeled individually and integrally to establish optimal influences through a structural model (Fig. 9).

The combined effects from all three components (proximal variables) are critical as well as the relation between them and the outcome (distal variables), which is within-year growth (on benchmark measures), to document change relative to peers. It is not enough to have only one of the proximal variables—the right measures, targeted interventions based on best evidence, and decisions tied to their effect on students. All three are needed. However, they need to work synergistically. And even then, changing these three components is not enough either.

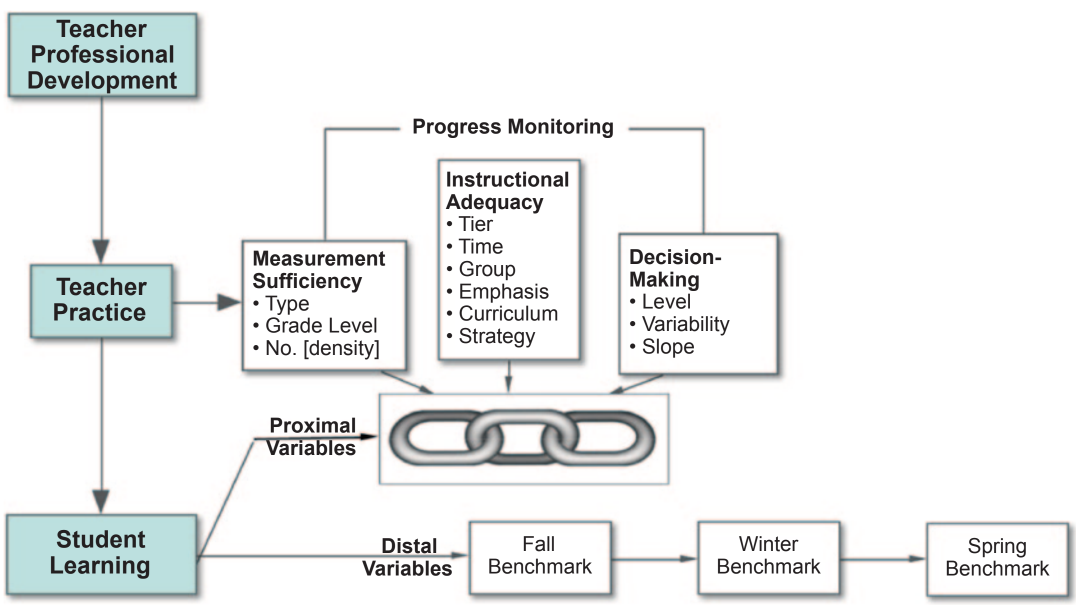


Fig. 9 CBM professional development → teacher change → student learning

Rather, the effect needs to close the achievement gap in which students at risk are catching up to their peers on grade-level performance measures (e.g., benchmarks). Finally, for systemic change, data need to be collected on proximal variables for developing reports on use, allowing for professional development to be tailored and specific.

Professional Development

Few studies have either documented training research on data-based decision-making (Stecker et al. 2005) or investigated effects of decision-making using a single-subject design (Hofstetter 2003). For the most part, training has focused on administration and scoring of measures rather than on how to use the outcomes to define critical components of instruction, integrating diagnosis of instruction based on decision-making that is hypothesis driven.

Stecker et al. (2005) documented investigations ($n=7$) that used "data-based decision rules" (with general outcomes) or used a skill analysis (mastery monitoring). The authors also considered in their summary whether achievement effects were found ($n=4$) or no achievement effects were found ($n=3$). In the end, they concluded training needed to move beyond simple collection of CBM data to affect student achievement. Rather, teachers must be trained to use the CBM measures to evaluate their instructional effectiveness; when student progress is less than expected, teachers then must be trained to make program modifications. They further noted that "raising goals when teachers underestimate student performance also appears to affect student growth. Data-management software also emerged as an aid to teachers in their use of CBM" (p. 803). Finally, when training included skills-analysis information, it "helped teachers examine student performance and highlighted skills for remediation" (p. 807). The authors concluded that not only do teachers need training but also the use of single-subject designs needs to be incorporated into the research on practice.

Hofstetter (2003) used a multiple-baseline (across passages) design with a first-grade student to investigate instruction and the addition of reward conditions to improve a student's oral reading fluency. Instruction included listening to the passage being read, repeated reading of the passage, and an error correction routine; in addition, a peer-mediated (tutoring) reading session was included. As the author reports, "The remarkable aspect of this finding is that the results were obtained with low-word-overlap passages" (p. 645). What this means is that when teachers are trained in a single-subject design, they can obtain significant results that transfer well beyond the corpus of words being directly taught. Importantly, single-subject design and methodology are the foundation upon which RTI rests, yet the RTI literature is less than rich in empirical examples and often this literature needs to be accessed through more traditional applied behavior analysis publications and search terms (e.g., *Journal of Applied Behavior Analysis* or *Journal of the Experimental Analysis of Behavior* using search phrases such as "brief experimental analysis" or "functional academic assessment"). In summary, an effective technology tool supporting RTI would include training in the systematic use of data, content standards covered, consideration of goals, skills analysis, and data management systems; it also would include use of single-subject designs in practice. However, most RTI training systems have little data on the effectiveness of the training, even though they are premised upon the collection of student performance and progress. Another problem arises when data are collected only at the individual student level and not on teachers. Therefore, professional development and practice cannot be tailored to the areas in which teachers need the most assistance, whether it is about how to effectively progress monitor students, how to develop effective instructional programs, or how to decide on maintaining or changing programs.

The graphic displays on technology-enhanced assessment systems can be designed to structure teachers' RTI, guiding them to make instructional decisions, while concurrently monitoring student progress. To be most effective for systems-wide application, such systems should also collect use data, so that policy-makers and administrators can apply the information to fine-tune their practice, improve the RTI system, and develop sensitive policies for practice. Presently, most RTI models operate primarily for individual students and on a laissez-faire basis with little feed forward or feedback in which progress is monitored, instruction is developed, or decisions are made. Although some research studies provide evidence of the effectiveness of using student performance data to assist teachers in providing interventions with integrity (see, e.g., VanDerHeyden et al. 2012), such checks and balances are not yet standard in assessment systems, providing rich ground for future development in this area.

The information on instruction through progress monitoring is expected to result in improved outcomes because the entire RTI process can become more systematic with evaluation data for teachers and students. For teachers, the feed forward and feedback features allow them to not only focus on the individual student but also to generalize in a manner that is likely to make their work more efficient. Training can focus on aggregated data within grade levels and measures and then be broken down in various ways within and across the three components of measurement sufficiency, instructional adequacy, and decision-making.

Technology-based assessments can provide teachers and students "next-generation" information through relational databases. For teachers, access can be provided to systematic models of assessment practices with classroom vignettes, exemplary practices, and resources that can be immediately used to develop effective progress monitoring. Student reports on content (standards) coverage as well as other aspects of instruction, combined with progress on both proximal and distal outcomes can be accessed to evaluate not only the effectiveness of interventions but also the systematic use of data-based decision-making. Such information can help teachers focus not only on how to implement best practice but also how to interpret information on student performance and progress. For example, a teacher could determine how many students are being monitored on specific measures aligned with specific standards, grade levels, and time intervals; how students are being organized into tiers, time, and groups, as well as specific instructional emphases being used (along with curriculum materials and strategies); how well decisions have been made with subsequent changes in level, variability, or slope. But realizing these potential benefits of technology-based assessments requires information and professional development in how to use the information in a systematic manner.

Enhancement to Training Through Effective Reports Future directions for enhancing technology-based assessment training, reports, and support materials (for easyCBM® and other learning management systems), should address the following components: measurement sufficiency, instructional adequacy, and decision-making.

- Information on benchmark performance for each of the measures should be analyzed and student performance used to recommend an appropriate measure type and grade level for progress monitoring (measurement sufficiency).
- Information on using the progress-monitoring system for specific students who are at risk and the number of days since the last progress measurement for each type and grade level of measure being used should be computed. This training could be used to select the student and the measure so the teacher can immediately begin measurement with either one-on-one administration (letter names and sounds, phoneme segmentation, and word or passage reading fluency) or group administration (vocabulary, comprehension, mathematics; measurement sufficiency).
- Use of benchmark measurement (and assignment to three levels of risk) can be used to group students by tier and size so teachers can organize groups with known levels of risk or ensure students at risk are distributed appropriately in various groups (instructional adequacy).
- Instruction/professional development should be embedded within the system, covering

such topics as: (a) instructional tier (1–3), (b) allocating time, (c) defining primary instructional emphasis (standards), (d) defining instructional strategy, and (e) providing curriculum and supplemental materials (instructional adequacy).

- Professional development should be included addressing using student progress to evaluate the current instructional program, including how to analyze for each progress measure (a) running average (cumulative level) of performance, (b) average variation (standard deviation), and (c) slope (displaying the average growth per week in the measure; decision-making).
- Annual goals and calculations the discrepancy between current changes and eventual (expected) performance using the three indices above with a "guesstimate" of likelihood (very likely, somewhat likely, and unlikely; decision-making).
- Information that informs decisions to "change or maintain" instruction associated with the number of measures (being used to calculate the progress) and the changeover (decision-making).

Approaching the topic of technology-based assessments by addressing warrants from a validity framework, a markedly different environment from current practice can be provided. First, the information can be integrated, using the power of databases to provide real-time reports drawing from a variety of sources (e.g., student performance data with individual, criterion, and normative referents; teacher instructional practices, including curriculum, strategies, intensity, duration, frequency, etc.; student demographics, enabling analysis of impact by various subgroups). Second, the system can be preventative, in that teachers may go back and forth between reports and actions. Third, such reports can tie together multiple databases for more appropriate reports on systems functioning targeted at professional development. This validity framework approach informs all parts of the discussion related to how the different parts of the technology-based assessment must work in concert to achieve the results sought in adopting an RTI approach: enhanced teacher understanding of data-based decision-making and improved student learning outcomes.

References

Alonzo, J., & Tindal, G. (2007). *The development of early literacy measures for use in a progress monitoring assessment system: Letter Names, Letter Sounds and Phoneme Segmenting (Technical Report No. 39)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2009a). *Alternate form and test-retest reliability of easyCBM® reading measures (Technical Report No. 0906)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2009c). *The development of K-8 progress monitoring measures in mathematics for use with the 2% and general education populations: Grade 1 (Technical Report No. 0919)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2009b). *The development of K-8 progress monitoring measures in mathematics for use with the 2% and general education populations: Kindergarten (Technical Report No. 0921)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM® ® online progress monitoring assessment system*. Eugene: University of Oregon, Behavioral Research and Teaching. (http://easyCBM®.com.).

Alonzo, J., Lai, C. F., & Tindal, G. (2009e). *The development of K-8 progress monitoring measures in mathematics for use with the 2% and general education populations: Grade 3 (Technical Report No. 0902)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Lai, C. F., & Tindal, G. (2009d). *The development of K-8 progress monitoring measures in mathematics for use with the 2 % and the general education populations: Grade 2 (Technical Report No. 0920)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Lai, C. F., & Tindal, G. (2009f). *Mathematics for Use with the 2% and General Education Populations: Grade 4 (Technical Report No. 0903)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Nese, J. F. T., Lai, C. F., Anderson, D., Park, B. J., & Tindal, G. (2010). *The alignment of easyCBM® math measures to curriculum standards (Technical Report No. 1002)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012a). *The development of CBM vocabulary measures: Grade 2 (Technical Report No. 1209)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012b). *The development of CBM vocabulary measures: Grade 3 (Technical Report No. 1210)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012c). *The development of CBM vocabulary measures: Grade 4 (Technical Report No. 1211)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012d). *The development of CBM vocabulary measures: Grade 5 (Technical Report No. 1212)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012e). *The development of CBM vocabulary measures: Grade 6 (Tehcnical Report No. 1213)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012f). *The development of CBM vocabulary measures: Grade 7 (Technical Report No. 1214)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012g). *The development of CBM vocabulary measures: Grade 8 (Technical Report No. 1215)*. Eugene: Behavioral Research and Teaching, University of Oregon.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anderson, D., Tindal, G., & Alonzo, J. (2009). *Internal consistency of general outcome measures in grades 1–8 (Technical Report No. 0915)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Lai, C. F., Nese, J. F. T., Park, B. J., Sáez, L., Jamgochian, E. M., & Tindal, G. (2010). *Technical adequacy of the easyCBM® primary-level mathematics measures (Grades k-2), 2009–2010 version (Technical Report No. 1006)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., & Tindal, G. (2010a). *easyCBM® mathematics criterion related validity evidence: Oregon state test (Technical Report No. 1011)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., & Tindal, G. (2010b). *easyCBM® mathematics criterion related validity evidence: Washington state test (Technical Report No. 1010)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Lai, C. F., Alonzo, J., & Tindal, G. (2011). Examining a grade-level math CBM designed for persistently low-performing students. *Educational Assessment, 16,* 15–34. doi:10.1080/10627197.2011.551084.

Anderson, D., Alonzo, J., & Tindal, G. (2011a). *A cross-validation of easyCBM® mathematics cut scores in Oregon: 2009–2010 (Technical Report No. 1104)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., & Tindal, G. (2011b). *A cross-validation of easyCBM® mathematics cut scores in Washington state: 2009–2010 Test (Technical Report No. 1105)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., & Tindal, G. (2011c). *easyCBM® reading criterion related to validity evidence: Washington state test 2009–2010 (Technical Report No. 1101)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., & Tindal, G. (2011d). *easyCBM® reading criterion related validity evidence: Oregon state test 2009–2010 (Technical Report No. 1103)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Lai, C., Park, B., Alonzo, J., & Tindal, G. (2012a). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM® reading assessments: Grade 2 (Technical Report # 1217)*. Eugene: Behavioral Research and Teaching.

Anderson, D., Park, B. J., Lai, C., Alonzo, J., & Tindal, G. (2012b). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM® reading assessments: Grade 1 (Technical Report # 1216)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum.

Hofstetter, C. (2003). Contextual and mathematics accommodation test effects for English-language learners. *Applied Measurement in Education, 16,* 159–188. doi:10.1207/S15324818AME1602_4.

Irvin, P. S., Park, B. J., Anderson, D., Alonzo, J., & Tindal, G. (2011). *A cross-validation of easyCBM® reading cut Scores in Washington: 2009–2010 (Technical Report No. 1109)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Irvin, P. S., Alonzo, J., Lai, C. F., Park, B. J., & Tindal, G. (2012a). *Analyzing the reliability of the easyCBM® reading comprehension measures: Grade 7 (Technical Report No. 1206)*. Eugegne: Behavioral Research and Teaching, University of Oregon.

Irvin, P. S., Alonzo, J., Park, B. J., Lai, C. F., & Tindal, G. (2012b). *Analyzing the reliability of the easyCBM® reading comprehension measures: Grade 6 (Technical Report No. 1205)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Alonzo, J., & Tindal, G. (2009a). *The development of K-8 progress monitoring measures in mathematics for use with the 2 % and general education populations: Grade 5 (Technical Report No. 0901)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Alonzo, J., & Tindal, G. (2009d). *The development of k-8 progress monitoring measures in mathematics for use with the 2 % and general education populations: Grade 8 (Technical Report No. 0904)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Alonzo, J., & Tindal, G. (2009c). *The development of k-8 progress monitoring measures in mathematics for use with the 2 % and general populations: Grade 7 (Technical Report No. 0908)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Alonzo, J., & Tindal, G. (2009b). *The development of K-8 progress monitoring measures in mathematics for use with the 2 % and the general education populations: Grade 6 (Technical Report No. 0907)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Nese, J. F. T., Jamgochian, E. M., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM® primary-level reading measures (grades K-1), 2009–2010 version (Technical Report No. 1003)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Irvin, P. S., Alonzo, J., Park, B. J., & Tindal, G. (2012a). *Analyzing the reliability of the easyCBM® reading comprehension measures: Grade 2 (Technical Report No. 1201)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Irvin, P. S., Park, B. J., Alonzo, J., & Tindal, G. (2012b). *Analyzing the reliability of the easyCBM® reading comprehension measures: Grade 3 (Technical Report No. 1202)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Liu, K., Ketterlin-Geller, L. R., Yovanoff, P., & Tindal, G. (2008). *Examining item functioning of math screening measures for grades 1–8 students (Technical Report No. 0804)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Nese, J. F. T., Lai, C. F., Anderson, D., Jamgochian, E. M., Kamata, A., Saez, L., & Tindal, G. (2010a). *Technical adequacy of the easyCBM® mathematics measures: Grades 3–8: 2009–2010 version (Technical Report No. 1007)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Nese, J. F. T., Lai, C. F., Anderson, D., Park, B. J., Tindal, G., & Alonzo, J. (2010b). *The alignment of easyCBM® math measures to curriculum standards (Technical Report No. 1002)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Park, B. J., Irvin, P. S., Lai, C. F., Alonzo, J., & Tindal, G. (2012a). *Analyzing the reliability of the easyCBM® reading comprehension measures: Grade 5 (Technical Report No. 1204)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Park, B. J., Irvin, P. S., Alonzo, J., Lai, C. F., & Tindal, G. (2012b). *Analyzing the reliability of the easyCBM® reading comprehension measures: Grade 4 (Technical Report No. 1203)*. Eugene: Behavioral Research and Teaching, University of Oregon.

Stecker, P., Fuchs, L., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795–819. doi:10.1002/pits.20113.

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments (Technical Report 42)*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. http://education.umn.edu/NCEO/OnlinePubs/Technical42.htm.

Tindal, G. (2013).Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *ISRN Education, 2013,* 1–29. doi:10.1155/2013/958530.

Tindal, G., Alonzo, J., & Anderson, D. (2009a). *Local normative data on easyCBM® ® reading and mathematics: Fall 2009 (Technical Report No, 0918)*. Eugene: Behavioral Research and Teaching, University of Oregon.

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Education Research Journal, 49,* 1251–1284.

Educational Technology and Response to Intervention: Affordances and Considerations

Janet S. Twyman and Melinda S. Sota

What is Educational Technology?

In the first half of the twentieth century, educational technology focused on media, as visual and then audiovisual tools were used to present instruction in forms such as film. This view of educational technology as hardware and software is still common today, and one might first think of educational technology in the form of computers and the educational software that runs on them (Reiser 2012).

Technology, however, involves application of research to solve practical problems and includes processes as well as tools (Clark and Salomon 1986; Twyman 2011). The notion of technology as a process became a focus of educational technology beginning in the 1950s, when educational technology came to be seen as involving the design of solutions for instructional problems and application of science to instructional practices (Reiser 2012).

In 2008, the Association for Educational Communications and Technology revised its definition of the field to include both resources and processes: "Educational technology is the study and ethical practice of facilitating learning and improving performance by creating, using, and managing appropriate technological processes and resources" (quoted in Reiser 2012, p. 4). In describing this definition, the authors define technological processes as "the systematic application of scientific or other organized knowledge to accomplish practical tasks" (quoted in Reiser 2012, p. 5), while technological resources refer to the hardware and software that we more typically think of when we think of educational technology (Reiser 2012).

Does Technology Impact Learning?

Whether and how technology impacts learning has been debated for several decades. The focus of this debate has been on whether the media used for instruction have unique effects on learning, apart from instructional methods. For example, Clark (1983, 1985, 1994) argued that media do not influence learning. Rather, it is the instructional method employed that influences learning, with the medium simply being a vehicle for a particular method. Kozma (1991, 1994) argued that different media might influence learning when they have different capabilities. Instructional methods could take advantage of these capabilities, and this interaction between a medium's capabilities and the methods that utilize them can result in more or different learning.

Any attempt to improve student learning must first stand on relevant, well-designed curricula and evidence-based instructional methods. Good

J. S. Twyman (✉)
University of Massachusetts Medical School/Shriver Center and The Center on Innovations in Learning, NY, US
e-mail: jstwyman@gmail.com

M. S. Sota
Illinois, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_29

instructional design requires a systematic design process that includes performing content, task, and learner analyses, clearly defining the learning objectives, determining the criterion tests to assess for understanding or mastery, establishing the entry repertoire needed by the student, building the instructional sequences, using performance data to continually adjust instruction, and ensuring student motivation by incorporating both program intrinsic and extrinsic consequences throughout the instructional sequence (Tiemann and Markle 1990; see also Dick and Carey 1996; Smith and Ragan 1999; Twyman et al. 2004). Good instructional delivery requires active learner engagement with frequent opportunities to respond (Rosenshine and Berliner 1978) and immediate, relevant, and contingent feedback (Bardwell 1981; Mory 1992; Shute 2008). Instruction should support the learner in moving forward at his or her own learning pace (Fox 2004) so that new material is not presented until the student has demonstrated mastery or application of current material (Bloom 1968; Keller 1968; Kulik et al. 1990). This requires that the progression of instruction and content be tied to actual measures of student learning, and not dictated by curriculum content chunks such as chapters or units, or the passage of marking periods or calendar years. Any viable educational technology must support, enhance, or provide these critical components.

Computer-based instruction (CBI) or computer-assisted instruction (CAI) has been the most prevalent form of hardware/software technology introduced into schools over the past half century. After initial fanfare with little tangible results (see AL-Bataineh and Brooks 2003), followed by more thorough empirical questions regarding the impact of CBI (see Shlechter 1991; Siegel 1994), the educational use of technology and CBI are gaining positive traction in the research literature. Improved outcomes have been demonstrated clearly in structured content areas such as mathematics (Valdez et al. 1999), social sciences (Kulik and Kulik 1991), and, with the growing involvement of the Internet and social connection in digital technology, contributions are beginning to be seen in other content areas as well (Redecker et al. 2010). Modern technology tools and applications such as video, interactive whiteboards, student response systems, portable devices, virtual learning, and a 1:1 ratio of computers to students have been found to greatly increase the collection, management, analysis, storage, and communication of educational data (McIntire 2002; Wayman 2005). Numerous meta-analyses of existing research have indicated a range of improvement effects for the use of computers, game-like curricula, and interactive simulations (Niemiec et al. 1987; Vogel et al. 2006). McNeil and Nelson (1990, cited in Hattie 2008) reported great variance in student outcomes due to factors such as instructional methods, learning materials, implementation variables, as well as the purpose(s) of the media. For example, Blanchard et al. (1999) conducted a meta-analysis of more than ten multimedia/game curriculum implementations for mathematics and language arts instruction in grades 1–5, and found overall low effects for mathematics ($d=0.13$) and language arts ($d=0.18$), yet higher effects ($d=0.23$) when interventions were implemented with quality. Vogel et al. (2006) analyzed studies that reported on differences in cognitive gains or attitudinal changes for computer games and simulations versus traditional classroom instruction. They found significantly higher cognitive gains for participants using games and simulations ($z=6.05$) as well as a main effect for attitude favoring the use of games and simulations ($z=13.74$).

A notable finding from more than two decades of computer-aided and multimedia interactive education is that increased "student control" over learning (such as pacing, sequencing, time allocation for mastery, choice of practice and review items) has resulted in equivocal outcomes compared to programs that were heavily or solely teacher directed (Niemiec et al. 1996). Some more recent studies have indicated that student attitudes towards school and subject matter (Roblyer 1989; Roblyer and Edwards 2000), as well as self-image and self-confidence (Alexiou-Ray et al. 2003; Christensen 2002; Roblyer et al. 1988), can be positively affected when using technology tools, although the durability of these effects is not known clearly.

Educational Technology and Response to Intervention

The purpose of response to intervention (RTI) is to "contribute to more meaningful identification of learning and behavioral problems, improve instructional quality, provide all students with the best opportunities to succeed in school, and assist with the identification of learning disabilities and other disabilities" (National Center on Response to Intervention, n.d., n.p.). A critical feature of RTI is the seamless integration of assessment and intervention to increase student learning and outcomes. RTI commonly employs three levels of intervention: primary prevention, secondary prevention, and tertiary prevention (Fuchs and Fuchs 2009). Primary prevention entails whole-class instruction delivered by a general education teacher. This instruction may or may not be differentiated and may or may not involve small-group and independent activities. The instructional materials used may or may not be empirically validated, but the design of most core programs is based on instructional principles. Secondary prevention involves additional small-group instruction using an evidence-based intervention protocol for those students who are identified as at risk. Reading or mathematics coaches may oversee this instruction, and paraprofessionals may deliver the instruction to small groups of students. Those students who do not make adequate progress in secondary prevention may move into tertiary prevention. This level employs intensive, individualized instruction (Fuchs and Fuchs 2009). The level of progress monitoring may also differentiate tiers, with more frequent progress monitoring often occurring with more intensive intervention (i.e., secondary and tertiary). RTI implementations may adjust the intensity and frequency of supplemental instruction and rate of progress monitoring to best meet the unique needs of the students receiving services, instructional demands of the classroom, and meaningful knowledge of a student's response to instruction.

Educational technology can help support RTI goals. The 2010 National Educational Technology Plan described five major goals for research and development related to educational technology (US Department of Education 2010). These five goals relate to (1) learning, (2) assessment, (3) teaching, (4) infrastructure, and (5) productivity, and are consistent with an RTI model. For example, the plan calls for research and development exploring how technologies such as simulations, virtual worlds, games, cognitive tutors, and collaboration environments can effectively motivate learners, while assessing complex skills and providing immediate performance feedback and adaptive instruction. Motivation, frequent assessment, feedback, and adaptive instruction are a few of the components that are part of an effective instructional model, and integration of assessment and intervention described within these goals also is a critical feature of RTI.

Technology also may help facilitate formative assessment and instructional decision-making. While research shows the effectiveness of RTI approaches in enhancing student learning (see Burns et al. 2005), fidelity of implementation is a critical variable in large-scale RTI implementation (Burns n.d.; Gansle and Noell 2007; Ysseldyke 2005). Technology may ease RTI implementation barriers, thus making RTI more likely to occur (Ysseldyke and McLeod 2007).

The promise of technology lies in its affordances. A teacher teaching a class of 30 students may ask an individual student a question, allowing that student a response opportunity. Other students in the class may or may not respond covertly to the question. If the teacher asks all students to respond—for example, by holding up response cards—it is more likely that all students will respond (Gardner et al. 1994; Narayan et al. 1990). If a teacher has a student response system, all students have an opportunity to respond, and the teacher can collect, analyze, and track individual and group responses over time, allowing for a more detailed assessment of student understanding and progress (Penuel et al. 2007). Technologies can provide teachers with a response system in which all students participating in instruction can respond, receive corrective feedback, and experience adjustments to instruction to further accelerate their learning. Supplemental instructional programs providing adaptive instruction and embedded assessment may also be used in all tiers—for example, as a supplement to core instruction in tier 1, for intervention in tier 2, and in a more focused way in tier 3 to help students gain fluency or fill skill gaps (Allsopp et al. 2010). The following sections describe some of the affordances technology can provide across tiers within an RTI model.

Technology in Tier 1 Instruction: Differentiating Instruction and Increasing Response Opportunities for All Learners

Many RTI theorists state that around 75–80 % of children achieve adequate levels of competency with the core curriculum alone (i.e., tier 1). Approximately 20 % of students are not successful in the core instruction despite good curriculum and generally effective instructional practices (Shapiro n.d.). Because most public schools in America do not have the resources to implement moderate-to-intensive intervention (i.e., tiers 2 and 3) to more than a quarter of the total student population, a strong core curriculum is foundational to student learning and a successful RTI implementation. Much work has been done in identifying evidence-based core curriculum programs and helping educators make informed decisions about what to use when teaching (especially in reading, see Foorman 2007; National Reading Panel—NRP 2000; Simmons and Kame'enui 2003).

Although instruction in tier 1 typically includes core instruction delivered to the whole class by the general education teacher, this instruction should be differentiated and include peer tutoring and flexible groupings (Lembke et al. 2012). However, in a review of the current state of RTI, Fuchs and Vaughn (2012) list differentiated tier 1 instruction as a continued challenge for classroom teachers. Although differentiated core instruction has been shown to reduce the number of students who require more intensive intervention and result in more proportionate representations of males, minorities, and English language learners in special education (Torgesen 2009; VanDerHeyden et al. 2007), this differentiation requires not only extensive knowledge of the subject matter but also the ability to use appropriate assessment tools to determine students' needs and effectively vary the type and intensity of instruction based on those needs. Improving instruction for primary prevention requires high instructional quality and opportunities to learn (Gerber 2005). Fuchs and Vaughn (2012) call for research and development of innovative instructional methods to improve tier 1 instruction. Given that the majority of instructional time is spent in a general education classroom, improving differentiated instruction in tier 1 has high potential for improving learning outcomes and decreasing the need for more intensive intervention.

Various forms of instructional technology have been found to influence and improve student learning in core curricula areas and general classroom instruction (Fadel and Lemke 2006; Flecknoe 2002; Gilbert 1996; Spector 2010). For example, student response systems have been found to improve student understanding and engagement, and create a more positive and interactive atmosphere (Caldwell 2007; Kay and LeSage 2009; Poole 2012). Reviews of research have generally agreed that the use of computers can increase student learning in a variety of subject areas and basic skills when combined with traditional instruction, and that students can learn more quickly and with greater retention when learning with computers. Student attitudes towards school and learning are also positively affected by the use of computers, and this use is most promising for at-risk and struggling learners (Fouts 2000).

Adaptive Instruction Across Tiers: Implementing Sound Instructional Methods and Progress Monitoring

Students who are struggling in an area also need explicit, systematic instruction with many opportunities for practice to build both accuracy and fluency. Instruction should also include cumulative and varied practice to promote retention and transfer of skills (Gersten et al. 2009). However, several challenges to implementing these instructional methods exist. For example, explicit instruction in mathematics may include providing a variety of models for problem-solving, verbalizing thought processes in teaching procedures and problem-solving methods, and offering guided practice and corrective feedback. Often, however, instructional materials include models of only one or two simple problems and may not include enough practice and cumulative review. (see, for example, Jitendra et al. 1996). Interventionists may also provide less practice than is necessary and may not be expert enough in the mathematics content to supplement a lack of modeling and practice within the materials (Gersten et al. 2009; Ma 1999).

Adaptive instruction within each tier is important in helping students make adequate progress and reducing the number of students who are moved to a more intensive tier. This can save time and resources, while still helping each student to succeed. However, using assessment data to make instructional decisions can be challenging for many schools, and skill in using data and implementing interventions may vary considerably among teachers (Kupzyk et al. 2012).

There are increasingly prevalent research-based, technology-enhanced interventions that assess and analyze current skills, target student deficits, and allow for automated instructional delivery. For example, Burst®: Reading by Wireless Generation® helps teachers continuously match reading interventions to each student's current ability and changing needs. Wireless Generation reports that smart technology allows learning data to be analyzed behind the scenes, recommends student groupings based on similar needs, and aligns instruction for the group. Similarly, Scholastic's *Read 180*® intervention program includes a "Groupinator" tool that suggests optimal small groups for differentiated instruction and then links those groups to appropriate resources. These software tools enhance implementation by automatically translating student data into specific intervention recommendations that teachers can then implement, either in their own teaching or by accessing other technology resources. Teachers can view what skills and concepts students have mastered, how much instruction was required for mastery, what areas might have caused particular difficulty, as well as the amount of time spent in instruction (or on a particular topic). Through analysis of the data, teachers can evaluate what educational materials produced the best outcomes and other behavioral and cognitive information relevant to academic performance over time. Such analytics can help educators determine the best instructional plan for groups of students, or any particular student (West 2011).

Motivation

Motivators are also important, as students who have historically had difficulty are less likely to engage in learning and practice opportunities as a result of frequent past failures (Fuchs et al. 2008). Motivators may be program extrinsic, such as awards, points, or badges for mastery or high levels of performance, and sites that purport to enhance student motivation through digital badges (such as Badgeville or Mozilla's Open Badges) or behavior management apps (such as Class Dojo), are just a few examples of motivational technology tools that may augment RTI. Motivators can also be program intrinsic. These motivators may arise from the instructional sequence and what mastery allows the learner to do in other contexts (Layng et al. 2004). For example, when an instructional sequence begins with a challenging task that a learner can do successfully, this experience of success may help the student more readily approach learning in that area (Fuchs et al. 2008). Computer-based programs that are able to continually assess and differentiate instruction (and therefore effectively align instruction with students' skill levels) may therefore promote student motivation by allowing for high rates of success in challenging tasks.

Games for learning have gained increasing attention in recent years, although games and "edutainment" have been part of educational technology for several decades. For example, the popular mathematics game Math Blaster® was first introduced in 1987. The structure of rewards in games may be especially effective in increasing motivation for struggling students, and can offer a learning environment in which feedback is less threatening (Shute 2008). Although games such as this are often thought of as offering practice in lower-level skills such as mathematical facts, games can also offer instruction and practice in higher-order thinking skills and problem-solving (Rice 2007).

Technology and Reading Intervention

Research Research has provided us with relatively clear guidelines about how to effectively teach children to read. In 2000, the NRP concluded a review of more than 100,000 studies that met scientifically based research standards and examined the effectiveness of an instructional approach in early reading that could be generalizable to a large number of students. They identified, based on an extensive body of knowledge, the skills children must learn in order to read well: phonemic awareness (the ability to manipulate individual sounds), phonics (the relationship between individual written letters and individual spoken sounds), fluency (the ability to accurately and quickly read text), vocabulary (the meaning of words), and comprehension (the understanding of what is being read). These skills often form the basis of not only core reading instruction but also the more focused work of a tier 2 or tier 3 reading intervention.

Until recently, relatively few studies have thoroughly evaluated new technologies for reading and literacy education. For example, Kamil and Lane (1998) reviewed the research in two mainstream literary journals with the highest citation rates for literacy research (*Reading Research Quarterly* and the *Journal of Reading Behavior,* since changed to *Journal of Literacy Research*) during the years between 1990 and 1995 and found that only 1 % of the articles discussed technology issues (also see Kamil et al. 2000). Within the past decade or more, there has been a growing number of examples of a technology assist for the five critical components of early reading instruction. Wise and Olson (1995) found that elementary students who received computer-assisted instruction in phonological awareness (by reading words in context and completing exercises involving individual words) made significant gains in phoneme awareness and word recognition. The use of screen-reading software (that converts text to digital speech) has helped improve comprehension, fluency, and accuracy and enhances concentration for special education students (Leong 1992; Lundberg and Olofsson 1993). Hearing a word spoken within the context of a passage helps students build decoding skills, word recognition, and vocabulary (Califee et al. 1991). Text to speech (TTS) software has been found to support comprehension by allowing the listener to focus on the meaning of the text without disturbing the text flow, thus increasing the ability to read interesting or grade-level materials, while minimizing the need for decoding skills (Wise et al. 2000).

Recommendations The US Department of Education's Institute of Education Sciences assists educators in identifying and implementing evidence-based interventions to increase student-reading achievement. In their What Works Clearinghouse report *Assisting Students Struggling with Reading: Response to Intervention (RTI) and Multi-Tier Intervention in the Primary Grades,* they identified five recommendations for effective RTI reading interventions:

- Screen all students for potential reading problems early and midway into the school year. Monitor the progress of students identified as "at risk" for developing reading disabilities.
- Provide evidence-based reading instruction, differentiated for all students based on assessments of current reading levels (tier 1).
- Provide intensive, systematic, and evidence-based instruction on foundational reading skills in smaller groups to students who score below "benchmark" (target score) on screening measures (tier 2).
- Monitor individual tier 2 student progress frequently (at least monthly) and use the data to determine which students still require intervention and may be in need of a tier 3 intervention plan.

- Provide daily, intensive, evidence-based instruction on the necessary components of reading proficiency to students who are far from the benchmark or show minimal progress after small-group instruction (tier 3).

Each of these recommendations can be supported with technology. Efficient, reliable screening measures and assessments may be provided online and evaluated automatically by the software. A database of reading screening scores over time allows for analysis at the student, class, or school level across years. Patterns in scores may allow for the identification of students, teachers, curricula, or systems that increase or decrease reading ability levels. Data analysis tools can help educators make data-driven decisions, or recommend differentiated instruction for students at varied reading proficiency levels. Learning management tools can help teachers plan and schedule time, instructional content, and degree of support and scaffolding based on student needs. Web-based clearinghouses or review sites can assist teachers in identifying various materials that support critical components of reading instruction (such as phonemic awareness, phonics, vocabulary, comprehension, and fluency), and that are prearranged to build skills gradually based on what has been previously taught. Technology programs can offer a high level of student interaction, providing precisely engineered learning opportunities and individualized feedback on responses, often to several students simultaneously. Technology programs can easily collect a student's response to each interaction, and parse those data based on important characteristics such as error patterns (to identify concepts that may require more instruction) or response latency (which may indicate the concept is not yet firm or fluent).

Within the domain of reading, technology assists can be grouped on various dimensions depending on purpose:

Fundamental Skills Programs, software, or apps that reinforce fundamental skills such as letter or phoneme identification, phonics, word attack, sentence construction, or symbol recognition. They may teach, review, or practice these types of skills in isolation or as part of a larger reading technology package.

Text-Reading/Text-to-Speech Software These programs convert written text into spoken words, with more modern technologies providing a more natural voice and cadence than available previously. Words may be highlighted in conjunction with their spoken output, or specific words may be selected for pronunciation or even definition.

Digital Text/Leveled Readers These resources allow teachers (and students) to identify books based on reading level or interest and may consist of classic literature, book summaries, study guides, picture books, or even interactive activities based on the material read. Many sites maintain a database of what has been read and provide "smart" recommendations (based on a user's individual choices and ratings), and often offer assistive technology enhancements like the ability to change text size, contrast, words per page, picture supports, and other enhancements.

Technology and Mathematics Intervention

Research Despite being a foundational skill critical for student success, the research base for effective mathematics instruction and intervention is not as extensive as that for reading (Fuchs et al. 2012; Lembke et al. 2012). For example, while reading studies have identified the critical component skills of phonics, phonemic awareness, fluency, comprehension, and vocabulary, component skills for mathematics have not been similarly identified. While it is possible that there are many more component skills required for mathematics (Fuchs et al. 2012), instruction in mathematics can be categorized in terms of three broad types of learning: conceptual, procedural, and strategic (Fuchs et al. 2008). For example, categorizing a story problem in terms of its type would be conceptual learning, carrying out the procedures to solve the problem would be considered procedural (Rittle-Johnson and Star 2009), and systematically attacking the problem—for example, starting by reading the problem carefully and ending by checking the work—would be considered strategic (Montague 1992; Polya 2004; Whimbey and Lochhead 1984).

Interventions utilizing technology have been shown to increase skills in mathematics. For example, in a review of studies investigating the effects of software programs on mathematics achievement, Kulik (2003) reported that out of 16 controlled studies conducted, 9 had an effect size large enough to be educationally meaningful. In all of the studies, test scores were at least slightly higher for the students engaged in the computer-based programs, and the median effect for all 16 studies was 0.38 standard deviations. Although technology used to increase fluency of basic fact recall is common, technology-based intervention programs can be effectively used for much more. For example, software programs may be particularly suited to delivering instruction in multiple levels of abstraction, such as moving from more concrete representations such as virtual manipulatives to abstract numerical representations, working with interactive simulations, and providing a variety of strategically delivered examples and nonexamples for conceptual learning. Technology can also be used in a supportive role, as when calculators are used to assist students in problem-solving when they may not be fluent in mathematics fact recall (Allsopp et al. 2010).

Recommendations In the primary grades, students who struggle with mathematics often have difficulty with number combinations (particularly in automatic retrieval of mathematical facts) and story problems. Also, although procedural instruction in mathematics is very common, conceptual instruction is often neglected (Fuchs et al. 2008). In the What Works Clearinghouse report *Assisting Students Struggling with Mathematics: Response to Intervention (RTI) for Elementary and Middle Schools,* eight recommendations were identified for effective RTI mathematics interventions:

1. Screen all students to identify those at risk for potential mathematics difficulties and provide interventions to students identified as at risk.
2. Focus on whole numbers in kindergarten through grade 5 and on rational numbers in grades 4–8.
3. Provide explicit and systematic instruction. This includes providing models of proficient problem-solving, verbalization of thought processes, guided practice, corrective feedback, and frequent cumulative review.
4. Include instruction on solving word problems based on common underlying structures.
5. Include opportunities for students to work with visual representations of mathematical ideas.
6. Devote about 10 min in each session to building fluent retrieval of basic arithmetic facts.
7. Monitor the progress of students receiving supplemental instruction and other students who are at risk.
8. Include motivational strategies in tier 2 and tier 3 interventions.

Each of these recommendations can be supported with technology.

Screening and Intervention Several programs are available for mathematical assessment and intervention, and software increasingly integrates assessment and instruction. For example, Wireless Generation®'s mCLASS® Mathematics formative assessment tool offers screening, diagnostic interviews, and progress monitoring tools as well as offering guidance on instructional interventions based on assessment results. Dreambox® Learning offers continuous adaptive instruction by tracking each mouse click within the program, using the data to identify student strategies, and adjusting instruction accordingly. As educational data mining and learning analytics continue to advance and grow more robust, the use of technology for continuous assessment and adaptive instruction will likely become more ubiquitous in educational technology products. (Bienkowski et al. 2012).

Explicit and Systematic Instruction Systematic instruction refers to the particular skills that are taught and the order in which they are taught, while explicit instruction refers to how those skills are taught (Kupzyk et al. 2012). Explicit and systematic instruction may include instructional techniques such as modeling, including think-aloud models of problem-solving, guided practice, corrective feedback, and frequent review (Gersten et al. 2009). Systematic instruction teaches component skills before those component skills are used in a more complex skill, building knowledge and skills in a logical order (Kupzyk et al. 2012). The sequence of instruction can also help to minimize learning challenges. For example, Fuchs et al. (2008) described an instructional sequence that began with what students already knew or could easily do for early success and then introduced new concepts and strategies as they became necessary and broadly applicable.

Several challenges exist to providing explicit and systematic instruction, and these are challenges that educational technology can help meet. For example, many instructional materials offer only a few models of problem-solving (Jitendra et al. 1996) and teachers or interventionists may not have the expertise in the subject matter necessary to provide additional models or talk through different strategies that could be used for problem-solving (Ma 1999). In addition, materials may lack appropriate levels of practice and review, particularly for students who are struggling (Gersten et al. 2009).

CBI programs that are developed based on a systematic and thorough analysis of the content and are able to analyze student errors can support teachers and interventionists by providing clear and varied models and carefully juxtaposed examples and nonexamples, can assess student strategy use and provide think-aloud models of the strategies. Finally, programs based on a mastery framework can provide practice and review based on learner performance, allowing those who have mastered the skills and strategies to move on, while providing more practice and review opportunities to those students who need them. It is important for companies developing educational software to take these instructional elements into account in the design of the program and for educators to evaluate potential software programs for these elements.

Instruction on Solving Word Problems Based on Common Underlying Structures Conceptual instruction is an important but often neglected aspect of mathematics instruction. Instruction should teach the underlying structure of different problem types, how to categorize problems based on their structure, and how to solve problems with a particular structure. However, instructional materials may not arrange instruction in a way that allows for classification of problem types and more complex problems are more difficult to classify (Gersten et al. 2009).

Although studies have shown the importance of conceptual instruction for word problems (see for example, Jitendra et al. 1998; Xin et al. 2005), conceptual instruction is not limited to word problems. For example, in her study comparing mathematics teachers' content knowledge in China and the USA, Ma (1999) described conceptual foundations of elementary mathematics. A purely procedural approach to subtracting two-digit numbers with regrouping would teach only the steps themselves, such as "borrowing" a ten from the tens column, adding ten ones to the ones column, and then subtracting. A conceptual approach, however, would teach fundamental concepts and principles that underlie the reasoning behind this algorithm, such as the meaning of place value and composing (and decomposing) a higher value unit.

Just as in providing explicit and systematic instruction, CBI programs can be developed to include instruction in a problem's underlying structure and practice categorizing and solving problems with different structures. In addition, programs can be developed to teach fundamental concepts and principles of mathematics and integrate conceptual and procedural instruction. However, it is important for companies developing programs to include this type of instruction and important for educators to look for these elements when evaluating educational software.

Opportunities For Students to Work With Visual Representations of Mathematical Ideas Mathematical ideas can be represented in a number of ways. Instructional programs should include work with concrete manipulatives, visual representations, as well as abstract symbols, and include consistent language across representations. However, working with different representations can be difficult in a classroom, and some interventionists may not have the content expertise to fully understand different representations—particularly for negative numbers, fractions, and proportional reasoning (Gersten et al. 2009). Ma (1999), for example, reported that most teachers she interviewed said they would use manipulatives in teaching subtraction with regrouping. However, when the teachers did not have strong content knowledge, the use of manipulatives was not directly related to the concept and therefore are not useful in teaching the skill. For example, two teachers suggested using counters such as beans in learning subtraction with regrouping. If the problem was 23−17, they would start with 23 beans and have students take away 17 beans. Although taking beans away illustrates subtraction, the students already understood subtraction, and the goal of the lesson is to teach the process of regrouping. Using manipulatives in the way that teachers described does not help students understand decomposing a higher-value unit in a base-ten system which is the key concept underlying regrouping. In fact, showing a child 23 beans and then asking the child to remove 17 to illustrate subtraction with regrouping makes no sense as an instructional strategy because it makes regrouping unnecessary altogether.

The visual and interactive nature of computer-based programs affords movement among representations and may be able to offer this type of instruction more easily and systematically than a teacher or interventionist. The National Library of Virtual Manipulatives (http://nlvm.usu.edu/en/nav/vlibrary.html) offers a variety of web-based virtual manipulatives in the form of Java applets, and virtual manipulatives are included in more extensive mathematics programs such as Dreambox® Learning.

Fluency Building Computer-based programs such as MathBlaster® have been offering opportunities for practice to build fluency for decades. Practice opportunities that are "gamified" can help to increase student motivation to practice. ExploreLearning's Reflex Math, for example, offers adaptive and individualized fluency practice with fact families in a game-playing context.

Limitations and Concerns of Technology-Based Interventions

Technology faces the same array of limitations that has plagued almost every innovation or major change impacting education. Factors such as an understanding of the purpose of the change, the need for properly trained staff, leadership and support for the change, and ongoing funding are fundamental for any successful change in schools. Ertmer (1999) grouped barriers to implementation into two broad categories: First-order barriers having to do with infrastructure such as access, time, training, support, and resources, and second-order barriers having more to do with the culture of the school and the individuals within it, such as attitudes, beliefs, practices, history of change, and resistance to change. These barriers and limitations, as well as potential solutions, are not endemic to an RTI technology implementation and have been comprehensively addressed elsewhere (see Barron et al. 2003; Earle 2002; Gülbahar 2007; Hope 1997; Leggett and Persichitte 1998; Lumley and Bailey 1993; Sheingold and Hadley 1990). Interested readers are encouraged to consult these resources.

Technology implementations may face additional barriers. Teaching (and learning) is viewed by many as a human, interpersonal endeavor, requiring an attention to the quality of the interaction and its attitudinal effects (O'Neal 1991). Technology is prevalent in every aspect of life, especially among youth, thus perhaps making learners more at ease with its use than teachers. Because technology is evolving so rapidly, knowledge and skills learned in one year may be obsolete in three to seven years, the amount of time research has shown it takes for an "implementation" to take hold and reap sustainable rewards (Fixsen et al. 2005). In addition to being facile in current technology uses, education policy-makers, curriculum specialists, technology specialists, school administrators, as well as those who develop technology products and services are required to "stay ahead of the curve" in order to properly prepare for changes ahead.

Perhaps the greatest overarching limitation to the successful use of technology in the classroom is educators' ability to find and effectively use technology to meet their teaching or their students' learning needs. The number of apps, tools, and resource sites, as well as commercial or enterprise technology programs from established educational publishers is huge, and continues to grow. In just under 2 years, from September 2009 to July 2011, the number of free apps in the Education category of Apple's ® ITunes® Store grew 369 %, from 866–3202. Paid educational apps grew by 202 %, from 4453 to 9013 (Gammon 2011). In 2012, almost three-quarters (72 %) of the top selling iTunes apps targeted preschool or elementary-aged children. Within the highly saturated games category, 32 % of apps stated an intended learning objective or made a claim of educational benefit (Shuler 2012). While the data are not readily available for the Android/Google app market, it can be expected that a similar growth trend is occurring. Teachers, curriculum specialists, technology specialists, and administrators must become "educated consumers" in the technology marketplace. With so many tools and programs available, sifting through the myriad of resources is a daunting task. Rubrics, guides, or checklists of necessary or notable characteristics of good technology can be helpful in determining what to use, when, and with whom. Such tools can help educators decide a technology's degree of:

- Relevance (Is there a strong connection between the learning goals or needs and the purpose of the technology?)
- Appropriateness (Does the technology fit the age, abilities, interest level of the learner—or educator?)
- Feedback (Does the technology let the user know when they are doing well, or provide additional help when needed?)
- Customization (Does the technology offer flexibility to alter content and settings to meet user needs?)
- Personalization (Does the technology adapt to learner needs—and interests?)
- Engagement (Does the technology increase efficient instructional time or capture learner interest?)
- Critical thinking (Does the technology encourage or support higher-order thinking skills including evaluating, analyzing, or creating?)
- Communication (Does the technology support the sharing of information or data?)

Appendices A–F provide examples of evaluation rubrics that may help educators determine the best use of certain technologies or tools. Rubrics may focus on different issues and should be carefully selected based on the context of the technology use. Some are relatively simple, like the "yes/no" checklist used in critical evaluation of an iPad/iPod App (Appendix A), while others present criteria aligned to Common Core Standards (e.g., Appendix D. Mobile Application Selection Rubric). Evaluate Apps for Special Needs, as the name implies, provides specific criteria for selecting apps to use with students with disabilities (see Appendix E). Evaluating the technology-learning environment is also important, as shown in the Arizona Technology Integration Matrix designed to help teachers to assess their own level of technology integration across learning environments (see Appendix F).

Issues for Future Research

Technology is constantly evolving, with new and innovative uses occurring all the time. The rate of technology change is accelerating exponentially (see Kurzweil 1999) across all areas of human endeavor, including education. Because of this rapid change, researchers have described their work with regard to educational technology as akin to chasing a "moving target" (Valdez et al. 1999, p. 1). Even as work is being conducted on the effects or implications of any particular educational technology, that technology itself is changing. While this evolving nature of educational technologies may hamper efforts to predict the success of, and establish guidelines for, subsequent educational practices (Leu 2000), many of the issues related to technology use remain constant. Factors such as the need for properly trained staff, sufficient equipment, and ongoing funding are essential for any successful integration of technology to increase learning.

Teachers, Technology, and Schooling Systems One of the most pivotal factors in the successful implementation of education technology is the teacher. Technology integration has moved beyond a handful of barely used, outdated computers in the back of the classroom or once or twice weekly class-wide forays into the computer lab to almost full-time integration of hardware, software, and Internet-access across all activities throughout the school day. When one considers that less than two decades ago barely half of our nation's teachers had Internet access at home (Becker et al. 1999), the technical savvy required of today's teachers might seem insurmountable. Expertise must surpass a basic understanding of hardware and software, and move into knowledge of the purposes of various software tools and how to use productivity tools, while following curriculum standards and adopting or maintaining a learner-centered perspective. Sang and colleagues (Sang et al. 2010) found that few teachers feel confident and competent in the goals and use of computer-based education in their classrooms. Even in classroom environments where technology is frequently used, researchers have found more emphasis on giving students access to information outside the classroom or on increasing student motivation, and less focus on how computers could improve specific academic achievement (Jostens Learning Corporation 1997) or be integrated with the curriculum or learning standards (Niess 1991; Trotter 1997).

The US Office of Technology Assessment reports one of the greatest roadblocks to integrating technology into a school's curriculum is the lack of teacher training, finding that most school districts spend less than 15% of their technology budgets on teacher training and development (US Congress, Office of Technology Assessment 1995). Ensuring understanding of pedagogical implications of technology integration is essential (Gilbert 1995; Watts and Hammons 2002). Instructional or educational technology should not be viewed as an add-on to teaching, but "integral to teaching practice" (Chism 2004, p. 43; see also Bates and Poole 2003; Grasha and Yanbarger-Hicks 2000) and critical to quality implementation (Shields and Behrman 2000). To be educated consumers of technology products, teachers must learn to select technologies that improve "the quality of teaching and learning [and] student motivation" (Gilbert 1996, p. 12), and research suggests that teachers who receive professional development focused on integrating technology into teaching may use the technology more effectively (Penuel et al. 2007).

In fact, all members of an educational ecosystem would benefit from increased understanding of effective uses of technology. "Digital media literacy continues its rise in importance as a key skill in every discipline and profession" (Johnson et al. 2011, p. 3). All educators need to be comfortable in evaluating technology resources for their students' and their own needs. They should be able to evaluate content, determine if it is culturally unbiased, current, appropriate to the curriculum standards, and respectful of student interest. Perhaps the most important quality is in understanding the educational goal before technology is selected or implemented, including a clear provision of how to seamlessly integrate the software into lesson strategies (Roblyer and Edwards 2000). As noted by Fullan (2000):

> Technology generates a glut of information, but it has no particular pedagogical wisdom—especially regarding new breakthroughs in cognitive science about how learners must construct their own meaning for deep understanding to occur. This means that teachers must become experts in pedagogical design. It also means that teachers must use the powers of technology, both in the classroom and in sharing with other teachers what they are learning. (p. 582)

Educational Technology Development Educational interventions and the companies who design and develop them should also be held to a high-quality standard in terms of student learning and engagement. As previously mentioned, good instructional design requires a systematic design process. That process should include iterative design and development with formative evaluation (Tiemann and Markle 1990; see also Dick and Carey 1996; Smith and Ragan 1999; Twyman et al. 2004). How a product is developed—whether it is developed based on best practices, evaluated for effectiveness after development (summative evaluation), or empirically tested during the design and development process (formative evaluation) should be considered when evaluating evidence of effectiveness (Twyman and Sota 2008).

Why Technology? What Technology Should Do.

By harnessing the power of digital and hardware advances merged with new knowledge and processes, we can further advance student learning and improve school outcomes. Using technology to assist teaching and learning may have started with stone carvings, papyrus, and the quill pen, progressed through the ages with the use of pencils, chalkboards, slide projectors, and TVs, and is now accelerating through the use of personal computers, laptops, tablets, and the power of the Internet and applications that leverage its reach and scale.

As noted, there is a strong literature base suggesting technology can improve instruction. However, the authors of this chapter and others (see Earle 1994; Fullan 2000; Rumph et al. 2007; Skinner 1968) suggest that it is not the "technology" (in this case hardware or software) itself that affects instruction, it is the philosophical underpinnings on which it is based *and* how it is used that influences its effectiveness. As noted by Wager (1992) "the educational technology that can make the biggest difference to schools and students is not the hardware, but the process of designing effective instruction" (p. 454). At least there is a fairly robust list of teaching and learning strategies with a strong evidence-base, across populations and subject matter, that have been shown to reliably improve learner outcomes (see Embry and Biglan 2008; Greer 2002; Lovitt 1994; Hattie 2008; Wolery et al. 1988). Any meaningful use of technology must support, augment, make easier, or make possible the myriad of things that we know empirically make a difference in children's lives. Any innovative use of technology must enable us to do important things that were not possible before.

Implications for Practice

At the District Level

Carefully consider, design, implement, and frequently review a district-wide technology plan:

- Improve organizational effectiveness by offering district-wide coordination and training to improve communication, planning, and record keeping.
- Purchase technology that supports greater efficiencies (i.e., doing more with less).
- Provide adequate equipment with plans for necessary upgrades.
- Plan for and support the integration of tools across sites.
- Plan for and support a common database.
- Plan for district-level interim assessments that support routine evaluation of instructional programs, and that provide credible, actionable data linked to relevant instructional resources.
- Align staff development with the district/school's technology goals.

At the School Level

- Encourage/support collaborative meetings by grade-level or subject matter to discuss planning and outcomes of technology-based instruction.
- Provide training and support in using student data and data systems to make instructional decisions.
- Conduct test runs of technology applications before widespread staff or student use (to identify any roadblocks or problems and avoid wasting valuable learning time).
- Plan for ongoing staff training to make effective use of the technology available.

- Provide professional development opportunities that are individualized to the teacher's level of expertise and experience and that focus on integrating technology into instruction.
- Identify model classrooms or peer mentors to allow other educators to see how various technologies can be integrated in teaching and learning.
- Provide peer coaching and mentor modeling to help the transition from knowing about (workshop information) to knowing how (classroom application and practice).
- Provide ongoing teacher support and opportunities for teachers to practice what they have learned (or to continue their learning).

At the Classroom/Teacher Level

- Use online tools that provide frequent or ongoing assessment to quickly understand what students know.
- Use frequent or ongoing measurement to tailor instruction to meet individual learning needs.
- Use active student response measurement systems to:
  - Check for real-time student understanding of content being taught
  - Display responses of the group and also occasion discussion and reflection
  - Gather formative data to guide instruction
  - Save time in administering and scoring quizzes
  - Incorporate individualized adaptive instructional programs as part of whole-class instruction as well as for intervention
- Use supplemental programs to provide additional practice opportunities.
- Consider using evidence-based educational games to increase student motivation and engagement.
- Select and use educational technology products for their affordances in terms of student interaction, engagement, and assessment (for example, use an interactive whiteboard to increase student interaction and not simply to present information).

At the Parent, Community Level

- Request openness and accountability, with verification of student benefit from expenditures.

At the Teacher/Administrator Pre-Service Level

- Emphasize the integration of technology into teaching, including offering courses on digital media pedagogy and literacy.

Appendix A. Critical Evaluation of an iPad/iPod App: Kathy Schrock

Name: ______________________________ Date:__________

CRITICAL EVALUATION OF AN CONTENT-BASED IPAD/IPOD APP

©2011-12. Kathleen Schrock (kathy@kathyschrock.net)
iPads in the Classroom site: http://linkyy.com/ipad

What is the title of the app? ______________________________ Cost: __________

Creator of the app ______________________ iTunes URL: ______________________

Content area(s): ______________________________ Grade level(s): ________

Content and components of the app	YES	NO
Curriculum connection: Are the skills reinforced connected to targeted skill/concept?	☐	☐
Authenticity: Are skills practiced in an authentic format/problem-based environment?	☐	☐
Feedback: Is feedback specific and result in improved student performance?	☐	☐
Differentiation: Does the app offers flexibility to alter settings to meet student needs?	☐	☐
User friendliness: Can students launch and navigate within the app independently?	☐	☐
Student motivation: Are students motivated to use the app and select it to use often?	☐	☐
Reporting: Is assessment/summary data available electronically to the student/teacher?	☐	☐
Sound: Does the music/sound in the app add to the educational aspects of the content?	☐	☐
Instructions: Are the instructions included within the app helpful to the student?	☐	☐
Support page: Does the app's supporting Web page provide additional useful information?	☐	☐

Level(s) of Bloom's Taxonomy addressed with this app (check all that apply)

☐ Remembering ☐ Understanding ☐ Applying ☐ Analyzing ☐ Evaluating ☐ Creating

Summary of the app								
Using the data you have collected above, explain why you would or would not recommend this application for use in the classroom. Include any specific ideas you have for its use.								
iTunes Application Ratings (click on ratings while in the App Store to learn more)	4+	☐	9+	☐	12+	☐	17+	☐

Adapted, with permission, from Harry Walker's "Evaluation rubric for iPad/iPad apps", 2010.

Appendix B. Educational App Evaluation Rubric: Tony Vincent

App Name: ______________________________ **Educational App Evaluation Rubric**

Purpose for App: ______________________________

	4	3	2	1
Relevance	The app's focus has a strong connection to the purpose for the app and appropriate for the student	The app's focus is related to the purpose for the app and mostly appropriate for the student	Limited connection to the purpose for the app and may not be appropriate for the student	Does not connect to the purpose for the app and not appropriate for the student
Customization	App offers complete flexibility to alter content and settings to meet student needs	App offers some flexibility to alter content and settings to meet student needs	App offers limited flexibility to adjust content and settings to meet student needs	App offers no flexibility to meet student needs
Feedback	Student is provided specific feedback	Student is provided feedback	Student is provided limited feedback	Student is not provided feedback
Thinking Skills	App encourages the use of higher order thinking skills including creating, evaluating, and analyzing	App facilitates the use of higher order thinking skills including evaluating, analyzing, and applying	App facilitates the use of mostly lower order thinking skills like understanding and remembering	App is limited to the use of lower order thinking skills like understanding and remembering
Usability	Student can launch and operate the app independently	Student needs to have a teacher show or model how to operate the app	Student needs to be cued each time the app is used	App is difficult to operate or crashes often
Engagement	Student is highly motivated to use the app	Student uses the app as directed by the teacher	Student perceives app as "more schoolwork" and may be off-task when directed to use the app	Student avoids the use of the app and might complain when its use is required
Sharing	Specific performance summary or student product is saved in app and can be exported to the teacher or for an audience	Performance data or student product is available in app but exporting is limited and may require a screenshot	Limited performance data or student product is not accessible	No performance summary or student product is saved

Rubric based on one edited by Kathy Schrock and originated by Harry Walker

Appendix C. Educational App Evaluation Checklist: Tony Vincent

Educational App Evaluation Checklist

App Name:

Purpose for App:

✓	
	Use of app is relevant to the purpose and student needs
	Help or tutorial is available in the app
	Content is appropriate for the student
	Information is error-free, factual, and reliable
	Content can be exported, copied, or printed
	App's settings and/or content can be customized
	Customized content can be transferred to other devices
	History is kept of student use of the app
	Design of app is functional and visually stimulating
	Student can exit app at any time without losing progress
	Works with accessibility options like VoiceOver and Speak Selection
	App is free of charge
	No in-app purchases are necessary for intended use of app
	App loads quickly and does not crash
	App contains no advertising
	App has been updated in the last 6 months
	App promotes creativity and imagination
	App provides opportunities to use higher order thinking skills
	App promotes collaboration and idea sharing
	App provides useful feedback
	Total ✓s *The more checks, the better the app is for education*

Educational App Evaluation Checklist

App Name:

Purpose for App:

✓	
	Use of app is relevant to the purpose and student needs
	Help or tutorial is available in the app
	Content is appropriate for the student
	Information is error-free, factual, and reliable
	Content can be exported, copied, or printed
	App's settings and/or content can be customized
	Customized content can be transferred to other devices
	History is kept of student use of the app
	Design of app is functional and visually stimulating
	Student can exit app at any time without losing progress
	Works with accessibility options like VoiceOver and Speak Selection
	App is free of charge
	No in-app purchases are necessary for intended use of app
	App loads quickly and does not crash
	App contains no advertising
	App has been updated in the last 6 months
	App promotes creativity and imagination
	App provides opportunities to use higher order thinking skills
	App promotes collaboration and idea sharing
	App provides useful feedback
	Total ✓s *The more checks, the better the app is for education*

This checklist is based on one originated by Palm Beach County Schools & Edudemic.com

Appendix D. Mobile Application Selection Rubric: eSkillsLearning

Mobile Application Selection Rubric

	WELL MEETS NEED	ADEQUATELY MEETS NEED	DOES NOT MEET NEED	COMMENTS
Aligned to Common Core Standards (CCS)	Documentation aligns app to CCS	App is loosely tied to CSS	App is not aligned to CSS	
Content presents in random order	Content is randomized; every game is new	App has several levels; Students will recognize game after play	App can be played once; same content repeated	
Engaging	Students will definitely like this	Students will probably like this	Students will not enjoy this	
Levels of difficulty	App will meet needs of all classroom groups	App has more than one level of difficulty	App has one level of difficulty	
Meets my students' needs	Students really need this content	App will be fun for students, but not generally needed	Students don't need this content	
Platform matches school equipment	App works with existing technology	App will work with some of school's technology	App isn't formatted to work with school's technology	
Research-based	Research-base is included with the app	Seems like there may be research to support this; I will need to find it	No research base for this app	
Scholastic presentation	App is very academic; supports serious learning	App isn't academic but has some learning value	App is a cute, fun game but little academic value	
Self-correcting	Feedback is provided; student either repeats work or is given instruction for learning	Feedback indicates right/wrong answer, then game proceeds	No feedback is provided	
Various modes of play	App allows for multiple players as well as play for different needs	App may be only used with a group	Single player only	

eSkills Learning

www.eskillslearning.net

P.O. Box 639
Candler, NC 28715
Toll Free (800) 638-6470

Appendix E. ievaluate Apps for Special Needs: Jeannette Van Houten

ievaluate app Rubric				
Goal: What goal from students IEP/504 does this app need to support?				
Name of App:			developer	
Content / Topic			developer Website:	
Date reviewed		Version:	Last up date:	
Review by:			Date:	Cost:
Domain	1 Weak Quality	2 Fair Quality	3 Good Quality	4 High Quality
Curriculum Connection	Does not meet expectation	Limited or narrow scope of the topic. Under developed.	Skills or concept are practiced and reinforced. Limited level of consideration.	Very strong connection to the skill or concept being practiced. Levels of consideration offered.
Type of Skills practices	No skill practice only "flashcard" drill	Skills are practiced in gaming format.	Simulated learning environment (virtual tasks). Scaffolds activities (Beginner - Advance)	Problem based learning with simulated environment. Program monitors and advances difficulty.
Age and Grade Level	Level is not appropriate for audience. Not suitable for age or grade level. Directions are incomplete or inadequate	Level is often too easy or difficult for target audience. Features unsuitable material. Directions are unclear.	Level is appropriate but some portions maybe to easy or difficult. Most directions are clear but some are confusing.	Level is appropriate for target audience (age and grade). Directions are clear and complete.
Languages	More than one language	2-3 languages	4-5 languages	6 or more languages
Adjustable levels	Only 1 level	2 -3 levels	4-5 levels	More than 5 levels
Prompts	No feedback offered moves forward with correct or incorrect responses	Prompt is limited to indicating wrong answer. Student needs to get it right to move forward	Prompt is specific - pre-set number of tries (can't edit) before student moves forward	Prompt is specific - can set number of tries - there is a tutorial to help student
Ease of Use	Very difficult to use. Limited or no instructions. Student needs support on every use	Student needs to be cued through the process.	Student needs support (model) from adult or another peer	Intuitive student can figure out independently
Engagement	Does not meet expectation	Held the individual attention for more than 2-3 minutes	Held the individual attention for more than 5 minutes	Held the individual attention for more than 10 minutes
Sub total				
Domain	1 Weak Quality	2 Fair Quality	3 Good Quality	4 High Quality
Customization	None	Can turn prompts off and music	Add your own items and prompts	All features are customizable including fonts.
Alternative Access	Has no access to alternative sources	Specific interface access and works consistently	App works with at least 2 access tools works consistenly	App works with 3 or more access tools. Is consistent.
Data Collected	No data offered.	Data is collected in percentage only. Data cannot be printed or stored.	Data is collected. Number of correct against total attempts. Can be printed.	Data is collected. Number of correct and incorrect responses against total attempts. Can be stored and printed.
National Curriculum	No			Yes
Gender Neutral	No			Yes
When was the app updated	a year or more	within the last 9 months	within the last 6 months	within the last 3 months
Sub total from this page 1		Sub total from this page		TOTAL______/14
Count the total of points divide by 14				
Rating:	Suitable for specific use	Satisfactory	Highly recommend	Exceeds expectations
Strengths of the APP:				
Weakness of the APP:				
Skills individual needs to have or learn before use:				
Alternative apps to consider				

Appendix F. AZ Technology Integration Matrix

TIM Arizona Technology Integration Matrix

What is the Arizona Technology Integration Matrix?

The Technology Integration Matrix (TIM) illustrates how teachers can use technology to enhance learning for K-12 students. The TIM incorporates five interdependent characteristics of meaningful learning environments: active, constructive, goal directed (i.e.reflective), authentic, and collaborative (Jonassen, Howland, Moore, & Marra, 2003). The TIM associates five levels of technology integration (i.e., entry, adoption, adaptation, infusion, and transformation) with each of the five characteristics of meaningful learning environments. Together, the five levels of technology integration and the five characteristics of meaningful learning environments create a matrix of 25 cells.

What is in each cell?

Within each cell of the Matrix, one will find two lesson plans with a short video of each lesson. Each lesson is designed to show the integration of technology in instruction and classrooms.

How should the Technology Integration Matrix be used?

The TIM is designed to assist schools and districts in evaluating the level of technology integration in classrooms and to provide teachers with models of how technology can be integrated into instruction in meaningful ways.

What is the history behind the tool?

The Technology Integration Matrix (TIM) was adapted from the Florida Technology Integration Matrix and developed through American Recovery and Reinvestment (ARRA) funds, which were administered by the Arizona Department of Education through the Pima County School Superintendent's Office. The goal of the Arizona TIM is to help provide a resource of technology integration in the classroom. Basic technology skills and integration of technology into the curriculum go hand-in-hand to form teacher technology literacy. Encouraging the seamless use of technology in all curriculum areas and promoting technology is essential in today's 21st Century Classroom. The Arizona TIM can help support the full integration of technology in Arizona's schools.

What are the next steps for developments with the Matrix?

We know that technology changes at a rapid pace. It is our intent that the TIM be a living document with additional lesson plans and videos added in the coming months and years. Districts and schools will be encouraged to use the TIM in the context of technology integration goal development and associated professional development planning. Through regular classroom observation and targeted professional development activities, it is our hope that, over time, teachers will be able to effectively monitor their progress through a continuum of technology integration levels.

The Arizona Technology Integration Matrix is adapted from the Technology Integration Matrix produced by the Florida Center for Instructional Technology. The Arizona Technology Integration Matrix was produced by the Arizona K12 Center at Northern Arizona University and funded in part through 2010-2011 American Recovery and Reinvestment Act (ARRA) funds administered by the Arizona Department of Education and the office of the Pima County School Superintendent in response to the *ARRA Enhancing Through Technology (EETT--Title IID) Ed Tech Standards Support Project.*

Appendix F. AZ Technology Integration Matrix, cont.

→ Levels of Technology Integration into the Curriculum

↓ Characteristics of the Learning Environment

Technology Integration Matrix	Entry	Adoption	Adaptation	Infusion	Transformation
	Teacher uses technology to deliver curriculum content to students.	Teacher directs students in the conventional use of tool-based software. If such software is available, this level is recommended.	Teacher encourages adaptation of tool-based software by allowing students to select and modify a tool to accomplish the task at hand.	Teacher consistently provides the infusion of technology tools with understanding, applying, analyzing, and evaluating learning tasks.	Teacher cultivates a rich learning environment, where blending choice of technology tools with student-initiated investigations, discussions, compositions, or projects, across any content area, is promoted.
Active Students are actively engaged in educational activities where technology is a transparent tool used to generate and accomplish objectives and learning.	Active: Entry Students receive content through the use of technology or use technology for drill and practice type activities.	Active: Adoption Students occasionally use specified technology tools to plan or create end products.	Active: Adaptation Students choose or modify the technology-related tools most appropriate for developing learning tasks.	Active: Infusion Students focus on learning tasks, and purposefully combine technology tools to design desired outcomes based on their own ideas.	Active: Transformation Students seamlessly organize the learning tasks and formulate products, discussions, or investigations using any appropriate technologies available.
Collaborative Students use technology tools to collaborate with others.	Collaborative: Entry Students primarily work alone in highly structured activities, using technology.	Collaborative: Adoption Students are allowed the opportunities to utilize collaborative tools in conventional ways.	Collaborative: Adaptation Students have opportunities to select and employ technology tools to facilitate and enhance collaborative work.	Collaborative: Infusion Students select technology tools to facilitate and enhance collaboration in all aspects of their learning.	Collaborative: Transformation Students seamlessly use technology tools to globally collaborate with peers and experts.
Constructive Students use technology to understand content and add meaning to their learning.	Constructive: Entry Technology used to deliver information to students.	Constructive: Adoption Students begin to use constructive technology tools to build upon prior knowledge and construct meaning.	Constructive: Adaptation Students have opportunities to choose and manipulate technology tools to assist them in molding their understanding.	Constructive: Infusion Students make connections with technology tools to construct deeper understanding across disciplines.	Constructive: Transformation Students use technology to construct, share, and publish new knowledge to an appropriate audience.
Authentic Students use technology tools to solve real-world problems meaningful to them, such as digital citizenship.	Authentic: Entry Students use technology to complete assigned activities that are generally unrelated to real-world problems.	Authentic: Adoption Students are allowed opportunities to employ technology tools to connect content-specific activities that are based on real-world problems.	Authentic: Adaptation Students have opportunities to select and utilize the appropriate technology tools and digital resources to solve problems based on real-world issues.	Authentic: Infusion Students select appropriate technology tools to complete authentic tasks across disciplines while modeling digital etiquette and responsible social interactions.	Authentic: Transformation Students participate in meaningful projects that require problem-solving strategies, and facilitate global awareness, through the utilization of technology tools.
Goal Directed Students use technology tools to research data, set goals, plan activities, monitor progress, and evaluate results.	Goal Directed: Entry Students receive directions, guidance, and feedback from technology, rather than using technology tools to set goals, plan activities, monitor	Goal Directed: Adoption From time to time, students have the opportunity to use technology to either plan, monitor, or evaluate an activity.	Goal Directed: Adaptation Students have opportunities to select and modify the use of technology tools to facilitate goal-setting, planning, monitoring, and/or evaluating specific activities.	Goal Directed: Infusion Students use technology tools to set goals, plan activities, monitor progress, and evaluate results throughout the curriculum.	Goal Directed: Transformation Students engage in ongoing metacognitive activities, with reflection or connected purpose, supported by technology tools.

References

AL-Bataineh, A., & Brooks, L. (2003). Challenges, advantages, and disadvantages of instructional technology in the community college classroom. *Community College Journal of Research and Practice, 27,* 473–484.

Alexiou-Ray, J., Wilson, E., Wright, V., & Peirano, A. M. (2003). Changing instructional practice: The impact of technology integration on students, parents, and school personnel. *Electronic Journal for the Integration of Technology in Education*. http://ejite.isu.edu/Volume2No2/AlexRay.htm.

Allsopp, D. H., McHatton, P. A., & Farmer, J. L. (2010). Technology, mathematics PS/RTI, and students with LD: What do we know, what have we tried, and what can we do to improve outcomes now and in the future? *Learning Disability Quarterly, 33,* 273–288.

Bardwell, R. (1981). Feedback: How does it function? *Journal of Experimental Education, 50,* 4–9.

Barron, A. E., Kemker, K., Harmes, C., & Kalaydjian, K. (2003). Large-scale research study on technology in K-12 schools: Technology integration as it relates to the national technology standards. *Journal of Research on Technology in Education, 35*(4), 489–507.

Bates, A. W., & Poole, G. (2003). *Effective teaching with technology in higher education*. San Francisco: Jossey-Bass.

Becker, H. J., Ravitz, J. L., & Wong, Y. T. (1999). *Teacher and teacher-directed student use of computers and software*. Teaching, learning, and computing: 1998 National Survey. Report #3.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: U.S. Department of Education, Office of Educational Technology.

Blanchard, J., Stock, W., & Marshall, J. (1999). Meta-analysis of research on a multimedia elementary school curriculum using personal and video-game computers. *Perceptual and Motor Skills, 88,* 329–336.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1–12.

Burns, M. K. (n.d.). Using technology to enhance RtI implementation. http://www.rtinetwork.org/getstarted/implement/using-technology-to-enhance-rti-implementation.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analysis of response-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE—Life Sciences Education, 6*(1), 9–20.

Califee, R., Chambliss, M., & Beretz, M. (1991). *Organizing for comprehension and composition*. In W. Ellis (Ed.), *All language and the creation of literacy* (pp. 79–93). Baltimore: International Dyslexia Association.

Chism, N. (2004). Using a framework to engage faculty in instructional technologies. *Educause Quarterly, 27*(2), 39–45.

Christensen, R. (2002). Effects of technology integration education on the attitudes of teachers and students. *Journal of Research on Technology in Education, 34*(4), 411–433.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 55*(4), 445–459.

Clark, R. E. (1985). Evidence for confounding in computer-based instruction studies: Analyzing the meta-analyses. *Educational Communications and Technology Journal, 33*(4), 249–262.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–29.

Clark, R. E., & Salomon, G. (1986). Media in teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 464–478). New York: Macmillan.

Dick, W., & Carey, L. (1996). *The systematic design of instruction* (4th ed.). New York: Harper Collins Publishing.

Earle, R. S. (1994). Instructional design and the classroom teacher: Looking back and moving ahead. *Educational Technology, 34*(3), 6–10.

Earle, R. S. (2002). The integration of instructional technology into public education: Promises and challenges. *ET Magazine, 42*(1), 5–13.

Embry D. D., & Biglan A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review, 11*(3), 75–113.

Ertmer, P. (1999). Addressing first- and second-order barriers to change: Strategies for technology implementation. *Educational Technology Research and Development, 47*(4), 47–61.

Fadel, C., & Lemke, C. (2006). Technology in schools: What the research says? http://www.cisco.com/web/strategy/docs/education/tech_in_schools_what_research_says.pdf.

Fixsen, D., Naoom, S. F., Blase, D. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Flecknoe, M. (2002). How can ICT help us to improve education? *Innovations in Education and Teaching International, 39*(4), 271–279.

Foorman, B. R. (2007). Primary prevention in classroom reading instruction. *Teaching Exceptional Children, 39,* 24–30.

Fouts, J. T. (2000). *Research on computers and education: Past, present, and future*. Seattle: Bill and Melinda Gates Foundation.

Fox, E. J. (2004). The personalized system of instruction: A flexible and effective approach to mastery learning. In D. J. Moran & R. W. Malott (Eds.), *Evidence based educational methods* (pp. 201–221). San Diego: Elsevier.

Fuchs, L. S., & Fuchs, D. (2009). On the importance of a unified model of Response-to-intervention. *Child Development Perspectives, 3*(1), 41–43.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195–203.

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31,* 79–92.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*(3), 257–269.

Fullan, M. (2000). The three stories of education reform. *Phi Delta Kappan, 81,* 581–584. Copyright (c) 2000.

Gammon, R. (2011, July 15). itunes app store educational apps 2011 vs. 2009. http://lh-llc.com/itunes-edu-apps-2011-2009.

Gansle, K. A., & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing resistance to intervention. In S. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 244–254). New York: Springer.

Gardner, R., Heward, W. L., & Grossi, T. A. (1994). Effects of response cards on student participation and academic achievement: A systematic replication with inner-city students during whole-class science instruction. *Journal of Applied Behavior Analysis, 27,* 63–71.

Gerber, M. M. (2005). Teachers are still the test: Limitations of response to instruction strategies for identifying children with learning disabilities. *Journal of Learning Disabilities, 38,* 516–524.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RTI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Gilbert, S. (1995). Technology and the changing academy. *Change, 27*(5), 58–61.

Gilbert, S. (1996). Making the most of a slow revolution. *Change, 28*(2), 10–23.

Grasha, A., & Yanbarger-Hicks, N. (2000). Integrating teaching styles and learning styles with instructional technology. *College Teaching, 48*(1), 2–10.

Greer, R. D. (2002). *Designing teaching strategies: An applied behavior analysis systems approach*. New York: Academic.

Gülbahar, Y. (2007). Technology planning: A roadmap to successful technology integration in schools. *Computers & Education, 49*(4), 943–956.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hope, W. C. (1997). Why technology has not realized its potential in schools? *American Secondary Education, 25*(4), 2–9.

Jitendra, A., Carnine, D., & Silbert, J. (1996). Descriptive analysis of fifth grade division instructions in basal mathematics programs: Violations of pedagogy. *Journal of Behavioral Education, 6*(4), 381–403.

Jitendra, A. K., Griffin, C. C., McGoey, K., Gardill, M. C., Bhat, P., & Riley, T. (1998). Effects of mathematical word problem solving by students at risk or with mild disabilities. *The Journal of Educational Research, 91*(6), 345–355.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Austin: The New Media Consortium.

Jostens Learning Corporation. (1997). *Survey analysis by global strategy group*. San Diego: Jostens Learning Corporation.

Kamil, M. L., & Lane, D. (1998). Researching the relationship between technology and literacy: An agenda for the 21st century. In D. Reinking, M. C. McKenna, L. D. Labbo, & R. D. Kieffer (Eds.), *Handbook of literacy and technology: Transformations in a post-typographic world* (pp. 323–341). Mahwah: Lawrence Erlbaum.

Kamil, M. L., Intrator, S. M., & Kim, H. S. (2000). The effects of other technologies on literacy and literacy learning. In Kamil M. L., Mosenthal P. B., Pearson P. D., & R. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 771–788). Mahwah: Lawrence Erlbaum.

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education, 53*(3), 819–827.

Keller, F. (1968). "Goodbye teacher …" *Journal of Applied Behavior Analysis, 1*, 79–89.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179–212.

Kozma, R. B. (1994) Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42*(2), 7–19.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say*. SRI Project Number P10446.001. Arlington: SRI International.

Kulik, C., & Kulik, J. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7,* 75–94.

Kulik, C., Kulik, J., & Bangert-Drowns, R. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*, 265–299.

Kupzyk, S., Daly, E. J., Ihlo, T., & Young, N. D. (2012). Modifying instruction within tiers in multitiered intervention programs. *Psychology in the Schools, 49*(3), 219–230.

Kurzweil, R. (1999). *The age of spiritual machines*. New York: Penguin Books.

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2004). Selected for success: How Headsprout Reading Basics™ teaches beginning reading? In D. J. Moran & R. Malott (Eds.), *Evidence-based educational methods*. St. Louis: Elsevier Science/Academic.

Leggett, W. P., & Persichitte, K. A. (1998). Blood, sweat, and TEARS: 50 years of technology implementation obstacles. *Tech Trends, 43*(3), 33–36.

Lembke, E. S., Hampton, D., & Beyers, S. J. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools, 49*(3), 257–272.

Leong, C. K. (1992). Enhancing reading comprehension with text-to-speech (DECtalk) computer system. *Reading and Writing: An Interdisciplinary Journal, 4,* 205–217.

Leu, D. J. (2000). Literacy and technology: Deictic consequences for literacy education in an information age. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 743–770). Mahwah: Lawrence Erlbaum.

Lovitt, T. C. (1994). *Tactics for teaching* (2nd ed.). Englewood Cliffs: Prentice-Hall.

Lumley, D., & Bailey, G. D. (1993). *Planning for technology: A guidebook for school administrators*. New York: Scholastic.

Lundberg, I., & Olofsson, A. (1993). Can computer speech support reading comprehension? *Computers in Human Behavior, 9,* 282–293.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah: Lawrence Erlbaum.

McIntire, T. (2002). The administrator's guide to data-driven decision making. *Technology & Learning, 22*(11), 18–33.

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities, 25*(4), 230–248.

Mory, E. H. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research and Development, 40*(3), 5–20.

Narayan, J. S., Heward, W. L., Gardner, R., Courson, F. H., & Omness, C. (1990). Using response cards to increase student participation in an elementary classroom. *Journal of Applied Behavior Analysis, 23,* 483–490.

National Center on Response to Intervention. (n.d.). What is RTI? http://www.rti4success.org/whatisrti.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read* (NIH Publication No. 00-4654). Bethesda: National Institute of Child Health and Human Development, National Institutes of Health.

Niemiec, R., Samson, G., Weinstein, T., & Walberg, H. J. (1987). The effects of computer based instruction in elementary schools: A quantitative synthesis [Abstract]. *Journal of Research on Computing in Education, 20*(2), 85–103.

Niemiec, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research, 15*(2), 157–174.

Niess, N. L. (1991). Computer-using teachers in a new decade. *Education and Computing, 7*(3–4), 151–156.

O'Neal, J. B., Jr. (1991). *Proceedings from frontiers in education twenty-first annual conference: Engineering education in a new world order*. Raleigh: North Carolina State University.

Penuel, W. R., Boscardin, C. K., Masyn, K., & Crawford, V. M. (2007). Teaching with student response systems in primary and secondary education settings: A survey study. *Educational Technology Research & Development, 55,* 315–346.

Polya, G. (2004). *How to solve it.* Princeton: Princeton University Press.

Poole, D. (2012). The impact of anonymous and assigned use of student response systems on student achievement. *Journal of Interactive Learning Research, 23*(2), 101–112.

Redecker, C., Ala-Mutka, K., & Punie, Y. (2010). Learning 2.0—The impact of social media on learning in Europe. Policy brief. JRC Scientific and Technical Report. EUR JRC56958 EN. http://www.ict-21.ch/com-ict/IMG/pdf/learning-2.0-EU-17pages-JRC56958.pdf.

Reiser, R. A. (2012). What field did you say you were in? Defining and naming our field. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed.). Boston: Pearson Education.

Rice, J. W. (2007). Assessing higher order thinking in video games. *Journal of Technology and Teacher Education, 15*(1), 87–100.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544.

Roblyer, M. D. (1989). *The impact of microcomputer-based instruction on teaching and learning. A review of recent research*. Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED346082).

Roblyer, M. D., & Edwards, J. (2000). *Integrating educational technology into teaching and learning* (2nd ed.). Upper Saddle River: Prentice-Hall.

Roblyer, M. D., Castine, W. H., & King, F. J. (1988). Assessing the impact of computer-based instruction: A review of recent research. *Computers in the Schools, 5,* 117–149.

Rosenshine, B. V., & Berliner, D. C. (1978). Academic engaged time. *British Journal of Teacher Education, 4*(1), 3–16.

Rumph, R., Ninness, C., McCuller, G., Holland, J., Ward, T., & Wilbourn, T. (2007). Stimulus change: Reinforcer or punisher? Reply to Hursh. *Behavior and Social Issues, 16*(1), 47–49.

Sang, G., Valcke, M., Braak, J. V., & Tondeur, J. (2010). Student teachers' thinking processes and ICT integration: Predictors of prospective teaching behaviors with educational technology. *Computers & Education, 54*(1), 103–112.

Shapiro, E. (n.d.). Tiered instruction and intervention in a response-to-intervention model. RTI Action Network.

http://www.rtinetwork.org/essential/tieredinstruction/tiered-instruction-and-intervention-rti-model.

Sheingold, K., & Hadley, M. (1990). *Accomplished teachers: Integrating computers into classroom practice*. New York: Bank Street College of Education, Center for Technology in Education.

Shields, M. K., & Behrman, R. E. (2000). Children and computer technology: Analysis and recommendations. *The Future of Children, Children and Computer Technology, 10*(2), 1–27.

Shlechter, T. M. (Ed.). (1991). *Problems and promises of computer-based training*. Norwood: Ablex Publishing.

Shuler, C. (2012). *iLearn II: Addendum, an analysis of the games category of the iTunes app store*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Siegel, J. (1994). No computer know how? *Electronic Learning, 13*(5), 58.

Simmons, D. C., & Kame'enui, K. J. (2003). A consumer's guide to evaluating a core reading program Grades K–3: A critical elements analysis. http://iris.peabody.vanderbilt.edu/rti03_reading/cons_guide_instr.pdf.

Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.

Smith, P., & Ragan, T. (1999). *Instructional design* (2nd ed.). New York: Wiley.

Spector, J. M. (2010). An overview of progress and problems in educational technology. *Interactive Educational Multimedia, 3,* 27–37.

Tiemann, P. W., & Markle, S. M. (1990). *Analyzing instructional content: A guide to instruction and evaluation*. Seattle: Morningside Press.

Torgesen, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in reading first schools. *Child Development Perspectives, 3*(1), 38–40.

Trotter, A. (1997). Taking technology's measure. In Technology counts: Schools and reform in the information age. *Education Week, 17*(11), 6–11.

Twyman, J. S. (2011). Emerging technologies and behavioural cusps: A new era for behaviour analysis? *European Journal of Behavior Analysis, 12*(2), 461–482.

Twyman, J. S., & Sota, M. (2008). Identifying research-based practices for RTI: Scientifically-based instruction. *Journal of Evidence-Based Practices for Schools, 9*(2), 86–97.

Twyman, J. S., Layng, T. V. J., Stikeleather, G., & Hobbins, K. A. (2004). A non-linear approach to curriculum design: The role of behavior analysis in building an effective reading program. In W. L. Heward et al. (Eds.), *Focus on behavior analysis in education, Vol. 3* (pp. 55–68). Upper Saddle River: Merrill/Prentice-Hall.

U.S. Congress, Office Of Technology Assessment. (1995). Teachers and technology: Making the connection. OTA-EHR-616 Washington DC: U.S. Government Printing Office. http://www.fas.org/ota/reports/9541.pdf.

U.S. Department of Education, Office of Educational Technology. (2010). *Transforming American education: Learning powered by technology*. Washington, DC: U.S. Department of Education, Office of Educational Technology.

Valdez, G., McNabb, M., Foertsch, M., Anderson, M., Hawkes, M., & Raack, L. (1999). *Computer-based technology and learning: Evolving uses and expectations*. Oakbrook: North Central Regional Educational Laboratory.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*, 225–256.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*, 229–243.

Wager, W. (1992) Educational technology: A broader vision. *Educational and Urban Society, 24*(4), 454–465.

Watts, G., & Hammons, J. (2002). Professional development: Setting the context. In G. E. Watts (Vol. Ed.), *Enhancing community colleges through professional development. New directions for community colleges* (number 120, pp. 5–10). San Francisco: Jossey-Bass.

Wayman, J. C. (2005). Involving teachers in data-driven decision-making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed At Risk, 10,* 295–308.

West, D. (2011). *Using technology to personalize learning and assess students in real-time*. Washington, DC: Brookings Institution Press.

Whimbey, A., & Lochhead, J. (1984). *Beyond problem solving and comprehension: An exploration of quantitative reasoning*. Philadelphia: The Franklin Institute Press.

Wise, B. W., & Olson, R. K. (1995). Computer-based phonological awareness and reading instruction. *Annals of Dyslexia, 45,* 99–122.

Wise, B., Ring, J., & Olson, K. (2000). Individual differences in gains from computer assisted-remedial reading. *Journal of Experimental Child Psychology, 77,* 197–235.

Wolery, M., Bailey, D. B., Jr., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional students*. Boston: Allyn and Bacon, Inc.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *The Journal of Special Education, 39*(3), 181–192.

Ysseldyke, J. E. (2005). Assessment and decision making for students with learning disabilities: What if this is as good as it gets? *Learning Disability Quarterly, 28,* 125–128.

Ysseldyke, J. E., & McCleod, S. (2007). Using technology tools to monitor response to intervention. In S. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 396–407). New York: Springer.

Response to Intervention for English Learners

Diana Socie and Mike Vanderwood

The quintessential response to intervention (RTI) concepts of prevention and early intervention are especially important for students who are English learners (ELs). ELs do not perform as well as their native English-speaking (NES) peers on large-scale accountability assessment systems (e.g., National Assessment of Educational Progress 2009). Yet evidence indicates that with the right type of educational opportunities, ELs can achieve levels of proficiency that are consistent with state and community expectations (Lesaux and Siegel 2003). Several authors argue that implementation of RTI can provide the appropriate support and experiences needed to lead to improved outcomes for ELs (Gersten et al. 2007; Vanderwood and Nam 2008).

In this chapter, we argue that teachers of EL students should implement an approach to RTI that combines a focus on literacy with English language development (ELD). In fact, we suggest ELD is best conceptualized as instruction that is provided in a multi-tiered approach that includes all the critical elements of an effective RTI approach for literacy (i.e., high-quality tier I instruction, screening, intervening, progress monitoring, and problem-solving). Literacy and ELD instruction are typically viewed as separate instructional objectives, yet literacy development is significantly dependent on the quality and intensity of the ELD support received by ELs. Mastery of academic language (i.e., the knowledge of the words necessary to access the curriculum) is arguably the single most important determinant of academic success for individual students (Francis et al. 2006). Oral language proficiency in English is crucial for text-level reading comprehension (August and Shanahan 2006).

To provide the reader with the context to understand why RTI should be applied with ELs, the chapter starts with a review of the statistics related to the growth of and need of support for ELs in US public schools. A very brief description of the connection between language development and reading is provided to emphasize the need to conceptualize reading concerns within language development. The next section emphasizes the need for high-quality core literacy and ELD instruction (i.e., tier I) combined with measures of those constructs that can be used to assess proficiency across the academic year (i.e., screening). After addressing tier 1, the chapter reviews current data on tier 2 interventions for reading and ELD, and highlights the lack of data-supporting tools for monitoring intervention effects (i.e., progress monitoring). Finally, the chapter ends with a set of recommendations about how to implement an integrated multitiered model that focuses on reading and ELD.

Two challenges are pertinent to making recommendations for evidence-based practices with ELs. First, research-validated techniques are in short supply. Second, the available research has

M. Vanderwood (✉) · D. Socie
University of California, Riverside, CA 92521, USA
e-mail: mike.vanderwood@ucr.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_30

limitations that reduce the utility of the information. For example, in many studies, authors treat ELs as a homogeneous group (Vanderwood and Nam 2008), yet it is fairly clear that English language proficiency for ELs moderates literacy outcomes and should be considered when making decisions about reading intervention focus (Gutierrez and Vanderwood 2013). In addition, the conclusions about what works for students whose first language is Spanish should not be applied to ELs who have a different first language. This chapter will emphasize high-quality, empirical research as the basis for practice recommendations. With the movement to prioritize high-quality experimental research and engage in evidence-based practices in schools (e.g., What Works Clearinghouse), the research basis for enhancing literacy skills among ELs is likely to continue to develop over the next decade.

Needs of ELs: Scope and Nature of the Problem

> If one considers the challenges facing English language learners who enter school with limited or no English at all, …the statistics are sobering. English language learners are "'school dependent" for English language development, and so the quantity and quality of exposure to rich and abundant language in school is absolutely essential. (White and Kim 2009, p. 5)

ELs represent one of the fastest-growing groups among school-aged children in the USA. Recent data indicate that approximately 21 % of children between the ages of 5 and 17 years (or 11.2 million children) speak a language other than English at home (Aud et al. 2011). In the past two decades, the population of ELs in schools has grown 169 %. This growth is in stark contrast to the general school population, which has grown only 12 % in the past 20 years. By 2015, it is projected that 30 % of the school-aged population in the USA will be ELs (Francis et al. 2006). The largest and fastest-growing populations of ELs in the USA include students who immigrated before kindergarten and US-born children of immigrants (Francis et al. 2006).

ELs speak more than 400 different languages with Spanish being spoken by the overwhelming majority, representing between 75 and 89 % of ELs (Office of English Language Acquisition or OELA 2012; Aud et al. 2011). While Spanish is the most common language spoken by ELs, states vary in terms of the linguistic diversity of their populations. In the 2006–2007 school year, Spanish was listed as the "most frequently spoken language" among EL students in 43 states and the District of Columbia, and 19 states reported that 80 % or more of their EL population were Spanish speakers (OELA 2012). However, the OELA (2012) reports that during the 2006–2007 and 2007–2008 school years, there were 12 states for which there was not one dominant language spoken by ELs. Table 1 includes the most recent analysis (i.e., 2007–2008) of the top five languages spoken nationwide by ELs.

Types of Schools Attended by ELs

According to recent data, Latino EL students, which make up the overwhelming majority of ELs nationwide, are significantly more likely than their Caucasian peers to come from a low socioeconomic background and to attend lower performing public schools (Fry 2008). Additionally, ELs are more likely to attend schools that are considered high poverty as measured by the

Table 1 Languages spoken in the USA

Language	Number of speakers	Percentage change from previous year
Spanish	3,767,749	0.75 % increase
Vietnamese	85,645	0.04 % decrease
Hmong	51,536	5.3 % increase
Arabic	41,557	6.4 % increase
Chinese	39,566	17 % increase

Note: Data taken from the Office of English Language Acquisition (2012) number of students who qualify for free or reduced-price lunch (Fry 2008). When reporting data, the federal government uses the percentage of students eligible for free or reduced-price lunch to classify schools as high- or low-poverty schools; schools where more than 75 % of students qualify for free or reduced-price lunch are considered high-poverty schools (Aud et al. 2011; NAEP 2009). Recent statistics from the US Department of Education indicate that in 2008–2009, there was an overwhelming overrepresentation of ELs in these high-poverty schools (NAEP 2009).

This overrepresentation in low-performing, high-poverty schools is of particular interest to researchers studying academic outcomes for ELs. According to a Pew Hispanic Center analysis, when ELs are not isolated in low-achieving schools, although there is still a performance gap, the gap in test score results is considerably narrower (Fry 2008). The Pew Hispanic Center proposes four major reasons why students in schools with high concentrations of ELs tend to fare worse than their counterparts in schools where the majority of students are Caucasian (Fry 2008): (1) schools with a high concentration of ELs tend to be located in city center or urban environments; (2) schools with a high concentration of ELs tend to have higher student enrollments; (3) there is a higher student-to-teacher ratio in schools, with a high concentration of ELs; and (4) students in schools with a high concentration of ELs tend to come from economically disadvantaged families.

Struggling EL Readers

Not only are ELs more likely to attend low-performing, high-poverty schools but they are also more likely than their native English-speaking peers to struggle to meet grade-level standards in all academic areas (Francis et al. 2006). Poor performance for ELs is especially evident in the area of reading, where recent data continue to highlight the achievement gap between Latino and Caucasian students (Aud et al. 2011). Since 2001 when states began routinely monitoring student acquisition of explicitly stated standards, researchers have been able to compare achievement between ELs and English-only students. Based on results from state standards testing in 2005, 73 % of fourth-grade Latinos (ELs and non-ELs) and 81 % of eighth-grade ELs scored in the below basic or well below basic categories in reading (Aud et al. 2011). That same year, only 25 % of Caucasian students in fourth grade and 19 % of Caucasian students in eighth grade obtained scores that were in the below basic or well below basic categories (Aud et al. 2011).

The newest information available from the National Assessment of Educational Progress (NAEP) through the Nation's Report Card (2011) indicates that while reading scores have improved overall since 2005, the achievement gap between Caucasian students and ELs has not changed. Additionally, in 2011, of all of the fourth graders nationwide who scored below the 25th percentile in reading, 24 % of them were ELs and 33 % were Caucasian, 25 % were Black, and 35 % were Hispanic/Latino. Conversely, of all of the fourth graders who scored above the 75th percentile, 2 % were ELs and 72 % were Caucasian, 7 % were Black, and 10 % were Hispanic/Latino.

The Language and Reading Connection

ELs face unique challenges learning to read in English. As previously discussed, many ELs come from low socioeconomic environments where access to robust language development is often limited. Hart and Risley (1995) determined that children who come from families where one parent is a welfare recipient enter kindergarten with less than half the amount of words in their vocabulary when compared to children who come from professional families (500 vs. 1100 words). Additionally, children who enter kindergarten discrepant in word knowledge continue to grow more discrepant each year of elementary school (Baker et al. 1997). However, there is evidence that children from linguistically diverse backgrounds can remediate this language gap if they have a strong foundation in their first language (Rhodes et al. 2005). Indeed, many scholars have theorized that languages develop interdependently, and that in the course of learning one language, children develop a set of skills and implicit metalinguistic knowledge that can be drawn upon when working in another language (Cummins 1984; Thomas and Collier 2002; Lesaux and Siegel 2003; Leafstedt and Gerber 2005). Cummin's (1979) *linguistic interdependence hypothesis* suggests the acquisition of a second language (e.g., English) in part depends on the adequate development of a student's native language (e.g., Spanish).

Many reading experts consider reading to be one aspect of language, such that reading problems can be considered as language problems (Fernald and Weisleder 2011; Robertson et al. 2012; Genesee and Geva 2006; Roth et al. 2002). Reading development occurs both simultaneously with and as a result of language development. Literacy in English requires that ELs must have thoroughly developed beginning reading skills, and must also comprehend English academic language (Cummins 1984; Solari and Gerber 2008; August and Shanahan 2006). Academic English includes understanding of semantic and syntactic language (Echevevarria et al. 2006), and is prerequisite for complex listening comprehension, vocabulary development, and reading comprehension (Echevevarria et al. 2006; Solari and Gerber 2008). Consequently, lack of established academic language is highly related to significant reading comprehension deficits (Echevarria et al. 2010; August and Shanahan 2006; Roth et al. 2002). Students who struggle with word meaning due to lack of language development are more prone to overall reading difficulties (Denton et al. 2008).

One area of language development that has been linked to reading outcomes is oral language development. Oral language skills are those skills that facilitate an understanding of semantics (e.g., comprehending and producing complex sentences), vocabulary, grammar, word order, and oral narration (Roth et al. 2002). English oral language proficiency is achieved through the systematic instruction of English vocabulary, listening comprehension, syntactic skills, and the awareness of the components of language (i.e., metalinguistic skills; August and Shanahan 2006; Francis et al. 2006). Few studies have focused on the impact of interventions for oral language development in the early grades, yet it appears this may be a promising area. In one longitudinal study looking at the connection between oral language and early reading, Roth and colleagues (2002) found that oral language skills of English-only students measured in kindergarten were better predictors than were phonological awareness skills of word and text-level reading performance at the end of second grade. They found that semantic language knowledge (i.e., word retrieval and oral language skills) in kindergarten was a robust indicator of text-level reading skills, and that meta-semantic skill (i.e., the ability to understand the meanings of words and sentences, including idioms, metaphors, and similes) measured in kindergarten contributed as equally to first-grade word reading as did phonological awareness skills (Roth et al. 2002). Although some studies have demonstrated that the development of early literacy skills in kindergarten and first grade were good predictors of word reading and text comprehension at the end of second grade (Lesaux and Siegel 2003), other studies have begun to demonstrate that English oral language proficiency is one of the key elements in ELs' development of word reading and text comprehension skills (August and Shanahan 2006; Roth et al. 2002).

ELD Instruction

Public schools are mandated to provide ELD instruction to students identified as ELs (formerly "limited English proficient" or LEP, students; Rhodes et al. 2005; Saunders and Goldenberg 2010). The goals of ELD instruction for ELs are (1) to ensure ELs acquire full proficiency in English as rapidly and effectively as possible and (2) to ensure that ELs, within a reasonable period of time, achieve the same rigorous grade-level academic standards that are expected of all students (NCELA 2011). Unfortunately, however, there is little existing research to help guide ELD instructional practices, and educators typically make decisions about what and how to teach ELD based on theory alone (Saunders and Goldenberg

Table 2 Descriptions of English language development programs

Program name	Description of program
Two-way immersion or two-way bilingual programs	The goal is to develop strong skills and proficiency in both L1 (native language) and L2 (English) Includes students with an English background and students from one other language background Instruction is in both languages, typically starting with smaller proportions of instruction in English, and gradually moving to half in each language Students typically stay in the program throughout elementary school
Dual language	When called "dual language immersion," usually the same as two-way immersion or two-way bilingual When called "dual language," may refer to students from one language group developing full literacy skills in two languages—L1 and English
Late exit transitional, developmental bilingual, or maintenance education	The goal is to develop some skills and proficiency in L1 and strong skills and proficiency in L2 (English) Instruction at lower grades is in L1, gradually transitioning to English; students typically transition into mainstream classrooms with their English-speaking peers Differences among the three programs focus on the degree of literacy students develop in the native language
Early exit transitional	The goal is to develop English skills as quickly as possible, without delaying learning of academic core content Instruction begins in L1, but rapidly moves to English; students typically are transitioned into mainstream classrooms with their English-speaking peers as soon as possible
Heritage language or indigenous language program	The goal is literacy in two languages Content taught in both languages, with teachers fluent in both languages Differences between the two programs: heritage language programs typically target students who are non-English speakers or who have weak literacy skills in L1; indigenous language programs support endangered minority languages in which students may have weak receptive and no productive skills—both programs often serve American Indian students
Sheltered English or Sheltered Instruction Observational Protocol (SIOP) Specially Designed Academic Instruction in English (SDAIE) Content-based English as a second anguage (ESL)	While there are some minor differences across these, the overall goal is proficiency in English while learning content in an all-English setting Students from various linguistic and cultural backgrounds can be in the same class Instruction is adapted to students' proficiency level and supplemented by gestures, visual aids May be used with other methods; e.g., early exit may use L1 for some classes and SDAIE for others
Structured English Immersion (SEI)	The goal is fluency in English, with only LEP students in the class All instruction is in English, adjusted to the proficiency level of students Subject matter is comprehensible Teachers need receptive skill in students' L1 and sheltered instructional techniques
English language development (ELD) or ESL pullout	The goal is fluency in English Students leave their mainstream classroom to spend part of the day receiving ESL instruction, often focused on grammar, vocabulary, and communication skills, not academic content There is typically no support for students' native languages

Note: For a more in-depth discussion on the purpose and structure of each program, please read Assessing Culturally and Linguistically Diverse Students (Rhodes et al. 2005).
LEP limited English proficiency

2010). Table 2 provides a description of several different methods of providing language instruction to ELs in public school settings.

Unlike their English-only peers, ELs face the specific challenge of acquiring English while simultaneously learning how to read and write in English (Francis et al. 2006). We now know that in order for ELs to succeed academically in the USA, they must acquire not only conversational skills in English but also academic language skills. Francis and colleagues (2006) stipulate that "lack of proficiency in academic language affects ELs' ability to comprehend and analyze texts in middle and high school, limits their ability to write and express themselves effectively, and can hinder their acquisition of academic content in all academic areas, including mathematics" (p. 7).

Unfortunately, ELs are not reaching proficiency in English as quickly as is necessary to achieve success in other content areas. Almost 60 % of English language learners in secondary schools have been in US schools for more than 6 years without reaching a sufficient level of English proficiency to be reclassified as fluent English speakers (Olsen 2010). Given that language development is a necessary component of academic development, these statistics exemplify the unique challenge that educators face in helping ELs become fluent readers.

As previously mentioned, schools are required to provide EL students with separate, ELD instruction to facilitate their mastery of academic English (Rhodes et al. 2005). There continues to be some debate in the literature around which types of programs lead to the best outcomes for EL students (Rhodes et al. 2005); however, the overwhelming majority of students in the USA obtain access to ELD through English-only immersion settings. In California, the state with the largest number of ELs, the California Department of Education (CDE 2012), identifies that of the approximately 1.4 million ELs in the state, more than 700,000 students are enrolled in Structured English Immersion settings, where students receive instruction in English only (Rhodes et al. 2005). It is important to note that English-only instruction with ELD support can take various forms, and may look differently depending on the programming choices each district makes. For example, in California, a total of 888,104 EL students receive at least one period of ELD and two periods of specially designed academic instruction in English (SDAIE) and/or sheltered instruction (SI) in subjects such as mathematics or social science in addition to the regular school offerings (CDE 2012).

Despite the vast number of students receiving ELD support through these English Immersion settings, it is clear that the achievement gap persists. Whether this problem is a result of the lack of empirically based, effective instructional strategies designed to help ELs acquire the English language remains to be determined. However, this problem is further complicated by issues related to the demographic and socioeconomic challenges associated with the majority of ELs, coupled with their lack of access to high-performing schools (see the section "Struggling EL Readers" for more information). Nonetheless, there is little debate that there is a need for high-quality reading and ELD instruction across schools, regardless of the location or performance level of the school.

Tier 1: The Need for High-Quality Reading and ELD Instruction

> There is no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum; for students who do not understand English are effectively foreclosed from any meaningful education. (Lau v. Nichols 1974)

Given the unique needs of ELs related to language development, educators must provide ELs with high-quality core instruction that focuses on providing instructional opportunities for developing academic language. Additionally, it is crucial that educators provide differentiated instruction for ELs to help develop their reading skills. This section begins with a description of one approach to providing structured academic content (i.e., tier I) instruction for ELs. This approach ensures that instructional practices are used that help to develop language while simultaneously giving students the necessary academic standards-based content. Next, a rationale is provided for using techniques that provide opportunities for differentiated reading instruction within core reading instruction. Specifically, two systems that have been established as effective practices for ELs and can be used with all students are reviewed. Finally, at the end of this section, research about reading screening with ELs is presented, and an argument is made for the need for separate norms and cut scores that account for English proficiency.

Instructional Practices that Support ELD

It is important here to distinguish between ELD and SI. Saunders and Goldenberg (2010) stipulate, "the primary goal of ELD instruction is learning and acquiring English" (p. 24). However, ELs must also be able to access content areas and require specific instruction in the academic English used in all content areas (e.g., English language arts, mathematics, science, social studies, and physical education). In other words, while both ELD and SI incorporate instructional strategies to help improve ELs' English language skills, SI specifically refers to strategies used to teach English knowledge in the content areas (Saunders and Goldenberg 2010). To further complicate this discussion, in some states SI is also referred to as SDAIE. Nonetheless, SI/SDAIE are instructional approaches that incorporate specialized strategies and techniques designed to increase ELs access to subject matter concepts, while simultaneously promoting students' ELD (Echevarria et al. 2008).

One approach to providing support to ELs during SI is through the use of the Sheltered Instruction Observation Protocol (SIOP; (Echevarria et al. 2008). While experts agree that the strategies utilized in SI are best practices for teaching ELs, at this stage there are very few studies demonstrating the effectiveness of these techniques. Researchers have had difficulty measuring the constructs and treatment integrity of SI practices and the strategies have historically been implemented inconsistently across districts due to the relative unavailability of professional development opportunities, instructional resources, and a comprehensive system of delivery (Echevarria et al. 2008; Echevarria et al. 2006). As a result, Echevarria and colleagues set out to develop a system of SI delivery that emphasized a standardized approach. This was the impetus for the development of the SIOP model, a lesson planning and delivery approach, for teaching ELs (Echevarria et al. 2006). Table 3 outlines the core components found in a lesson designed with the SIOP model.

A number of studies have compared student outcomes given tier 1 ELD support through the SIOP model of instruction. Many of these studies have compared literacy outcomes for students receiving instruction from teachers who had received different levels of training of the SIOP model. Additionally, studies have compared outcomes for students whose teachers utilized the SIOP model of instruction versus students whose teachers used basic SI practices in delivering instruction (Echevarria et al. 2008). Across studies, the research indicates that students who received instruction through the SIOP model, from teachers who were highly trained in the instructional practices and implemented procedures with fidelity, outperformed their peers on the outcome measures (Echevarria et al. 2008; Echevarria et al. 2006; Short et al. 2011).

The model was developed first as an observation tool to help researchers measure the treatment fidelity of SI delivery (Short et al. 2011). However, after 4 years of observation, Short and colleagues began development of a lesson plan and instructional delivery tool to improve outcomes for ELs. In a quasi-experimental study, Short and colleagues set out to determine to what extent the SIOP instructional delivery system improved outcomes in reading, writing, and oral language for ELs.

The study was conducted in the eastern USA in two matched districts (one treatment and one control). A total of 580 students participated with 387 students in the treatment condition speaking more than 15 languages, and 193 students in the control condition speaking eight different languages. In the treatment district, secondary school teachers (grades 6–12) were provided with

Table 3 Eight components of the SIOP model of instruction. (Echevarria et al. (2010))

Component	Description
Lesson preparation	The features under lesson preparation examine the lesson planning process, including the incorporation of language and content objectives linked to curriculum standards. In this way, students gain important experience with key grade-level content and skills as they progress toward fluency in the second language. Other features include the use of supplementary materials and meaningful activities
Building background	Building background focuses on making connections with students' background experiences and prior learning, and developing their academic vocabulary. The SIOP model underscores the importance of building a broad vocabulary base for students to be effective readers, writers, speakers, and listeners. In the SIOP model, teachers directly teach key vocabulary and word structures, word families, and word relations
Comprehensible input	Comprehensible input considers adjusting teacher speech, modeling academic tasks, and using multimodal techniques to enhance comprehension (e.g., gestures, pictures, graphic organizers, restating, repeating, reducing the speed of the teacher's presentation, previewing important information, and hands-on activities). The academic tasks must be explained clearly, both orally and in writing, with models and examples of good work so students know the steps they should take and can envision the desired result
Strategies	The strategies component emphasizes explicit teaching of learning strategies to students so that they know how to access and retain information. Good reading comprehension strategies, for example, need to be modeled and practices, one at a time with authentic text. SIOP teachers must scaffold instruction so students can be successful, beginning at the students' performance level and providing support to move them to a higher level of understanding and accomplishment. Teachers have to ask critical thinking questions as well so that students apply their language skills while developing a deeper understanding of the subjects
Interaction	Interaction features encourage elaborate speech and grouping students appropriately for language and content development. They need oral language practice to help develop content knowledge and second-language literacy; thus, student–student interaction is important and needs to occur regularly in each lesson. ELs need to practice important language functions, such as confirming information, elaborating on one's own or another's idea, and evaluating opinions
Practice/application	Practice/application calls for activities that extend language and content learning by encouraging students to practice and apply the content they are learning, as well as their language skills. It is important to build and reinforce reading, writing, listening, and speaking skills within content learning
Lesson delivery	Lesson delivery ensures that teachers present a lesson that meets the planned objectives. Successful delivery of an SIOP lesson means that the content and language objectives were met, the pacing was appropriate, and the students had a high level of engagement
Review/assessment	English language learners need to revisit key vocabulary and concepts, and teachers need to use frequent comprehension checks throughout lessons as well as other informal assessment to measure how well students understand and retain the information. Each SIOP lesson should wrap up with some time for revise and assessment and time to determine whether the lesson's objectives were met

SIOP Sheltered English or Sheltered Instruction Observational Protocol

7 days of training and professional development on how to use the SIOP model of instruction. Additionally, there were on-site coaches in all of the schools to assist with implementation. Teachers in the control condition were not provided with any specific SIOP training support; however, some of the teachers used SI programming with their EL population. In order to assess the extent to which teachers in the control district used SIOP instructional strategies, the researchers observed teachers on various occasions. They found that 5 % of teachers in the control condition implemented the SIOP to a high level, compared to 71 % of teachers in the treatment district.

The design of the study incorporated pre- and posttest measures for both groups of students that consisted of the International Dialects of English Archive (IDEA) language proficiency tests (the state standardized assessment of English language proficiency) and student performance on several state content achievement tests. They found that, by the end of the second year of implementation, the students' mean scores in the treatment district were significantly higher on both measures than those in the control district. The effect size (Cohen's *d*) for the treatment condition was calculated at 0.29 for the language proficiency tests, 0.16 for reading, 0.31 for writing, and 0.23 for total English proficiency.

It is important to note here, however, that there were some limitations to this study that Short and colleagues have attempted to address in subsequent studies. The first limitation has to do with the characteristics of the control versus the treatment group of students. The researchers did not use a randomized controlled trial to control for nonmeasured differences in the groups, and did not use a systematic approach to match students in the treatment and control condition. In other words, the students in the control condition spoke fewer languages and comprised different nationalities than those students in the treatment condition. They also mention that since they used students in secondary school, they hypothesized that they saw lower effect sizes than they would have had they conducted the study with elementary school children.

In a follow-up study, Short and colleagues employed a randomized controlled trial with a total of 1021 students (649 in the treatment condition and 349 in the control condition) in ten middle schools in Southern California (Short et al. 2011). In this study, they also included students who were not ELs for the purposes of determining if the SIOP system could help improve the academic language of native English speakers (NES) as well as ELs. The SIOP system was used during seventh-grade life science classes over one semester in the treatment schools. The outcome variables were a pre- and post-CREATE (Center for Research on the Educational Achievement and Teaching of English Language Learners) science language assessment (essay) and a science content measure developed by the district.

Results indicated that students in the treatment condition outperformed students in the control condition, but not to a statistically significant degree. Effect sizes were calculated using a pooled within-groups estimate of the standard deviation (SD), and were estimated to be small ($g=0.103$ for the nonessay portion of the posttest, and $g=0.197$ for the essay portion of the posttest). The authors reported that they were not surprised by the results given that few teachers implemented the SIOP model with fidelity. When they compared the outcomes between weak and strong implementers of the SIOP model, they found that students whose teachers implemented the model to a strong degree outperformed those students whose teachers did not. These results were statistically significant at $R^2=0.22, p<0.05$. Thus, treatment fidelity was a limitation in this study. Additionally, the use of a test without reported reliability and validity for the outcome measure served as a threat to statistical conclusion validity (Shadish et al. 2002). (For a list of all of the studies, please visit http://siop.pearson.com/about-siop/research.html.)

Differentiated Reading Instruction

In addition to providing explicit instruction in ELD with a particular focus on English academic language, effective reading instruction must also be provided. Differentiated instruction is the practice of matching instruction using evidence-based practices to meet the differing needs of learners in a given classroom (Kosanovich et al. 2005; Jones et al. 2012). It has been conceptualized as a key component of tier 1 practices in an RTI model, where effective teaching strategies can be maximized to provide support for diverse learners in the classroom (Jones et al. 2012). Differentiation can be achieved through a variety of instructional strategies, including the use of small groups directed by a teacher or aide, reading centers, or reciprocal peer teaching (Kosanovich et al. 2005; Saenz et al. 2005). While there are various ways teachers can differentiate instruction in a classroom, two reciprocal peer teaching strategies in particular have been found to be promising interventions with ELs: Peer-Assisted Learning Strategies (PALS; Saenz et al. 2005; visit http://kc.vanderbilt.edu/pals/ for information on ordering and further research) and Collaborative Strategic Reading (CSR; Klinger and Vaughn 1999; visit http://www.soprislearning.com/literacy/collaborative-strategic-reading for information on ordering).

PALS is a reciprocal classwide peer-tutoring strategy which has different goals for implementation in kindergarten and first grade than it does for second through sixth grade (Saenz et al. 2005). In kindergarten and first grade, teachers can use PALS to build beginning reading skills, whereas in second through sixth grade, the focus is on increasing strategic reading behavior, reading fluency, and comprehension (Saenz et al. 2005; Calhoon et al. 2007). For the second- through sixth-grade intervention specifically, PALS consists of students working together in pairs to deconstruct the meaning of a passage by use of partner reading with story retell, "paragraph shrinking" (i.e., students stopping every 5 min after reading aloud with a partner and summarizing what was just read), and engaging in prediction of upcoming text (Saenz et al. 2005). Additionally, during PALS, students engage in reading aloud for a significant portion of the intervention time on passages that have been matched to their level of proficiency by their teacher with guided practice provided by their partners. While PALS has been determined to be useful for students of all ability levels, Saenz and colleagues (2005) list five reasons why PALS is particularly suited for use with ELs: (1) PALS provides frequent opportunities for students to practice language related to a passage, (2) PALS requires that students engage in higher-order language activities when summarizing and making predictions, (3) PALS allows for individualization of support in reading through ability-appropriate reading passages, (4) PALS provides an opportunity for students to correct themselves and receive correction from other students when they say something incorrectly, and gives them opportunities to practice the correction, and (5) through its use of pairs, PALS helps facilitate development and use of English language skills in the classroom.

In a recent study, Saenz and colleagues implemented PALS in the third through sixth grade in one school district in the southwestern USA. Saenz et al. (2005) randomly assigned bilingual education classrooms to either a PALS treatment condition or a control no-treatment condition. In the no-treatment condition, teachers continued to teach reading as they previously had with no changes to the curriculum or instructional groupings within the classroom (Saenz et al. 2005). In total, 132 native Spanish-speaking students, including students with learning disabilities, and low-, average-, and high-achieving students, received the PALS intervention (Saenz et al. 2005). In this study, students engaged in PALS activities three times per week for 35 min per session over 15 weeks during their regularly scheduled reading instruction.

Results indicated that EL students who participated in PALS, with and without learning disabilities, made statistically significant gains in reading comprehension when compared to the students in the no-treatment control condition (Saenz et al. 2005). The authors calculated the effect sizes by calculating the difference between mean improvement scores divided by the SD. They reported the largest effect size for students with learning disabilities on improved number of words read on the Comprehensive Reading Assessment Battery (CRAB; $ES=1.01$). Despite the fact that there was not a statistically significant result in this area, overall, they reported that the average effect size for all students for improvement on number of words read was 0.60 (Saenz et al. 2005). With respect to improvement in the number of questions answered correctly, again, the students with learning disabilities had larger effect sizes than the other students ($ES=1.03$), with the average effect size across students reported as 1.02 (Saenz et al. 2005). It is important to note that the results of this study were statistically significant only for reading comprehension. The use of transitional bilingual education classrooms may limit the external validity of the findings to students who are in a sheltered immersion setting for ELs.

Similar to PALS, CSR is an empirically supported approach for differentiating instruction for culturally and linguistically diverse (CLD) learners in grades 4 through 12 (Klinger and Vaughn 1999; Vaughn et al. 2001; Vaughn et al. 2000). In CSR, students of differing ability levels work in small groups in the classroom to help each other achieve success in four domains of content area text: (1) making predictions prior to reading (i.e., preview), (2) identifying difficult words and concepts (i.e., click and clunk), (3) stating the most important parts in a section of text (i.e., get the gist), and (4) summarizing the reading (i.e., wrapping up; Klinger and Vaughn 1999). CSR is beneficial for ELs, because it provides an instructional practice that enhances comprehension of text and a formalized procedure that facilitates peer interaction and peer-mediated instruction (Klinger and Vaughn 1999). Vaughn and colleagues have found that implementation of CSR in classrooms with ELs has increased ELs academic engaged time, academic-related strategic discussion with peers, reading comprehension, and overall reading outcomes (Vaughn et al. 2001).

During the development phase of CSR, Klinger and Vaughn (1996) examined the effects of using CSR with ELs with a modified version of the CSR that is used today. They implemented reciprocal teaching strategies in groups of eight students for 15 sessions. The students who received the treatment were 26 Latino EL middle school students with learning disabilities. After the 15 sessions, those 26 students were then randomly assigned to one of two conditions—either tutoring younger students with LD or working in small groups on comprehension strategies without a teacher (Klinger and Vaughn 1996). Despite the small number of participants, Vaughn and colleagues reported a statistically significant gain in reading comprehension for students in both conditions. Effect sizes were not reported for this study.

Since the initial study in 1996, various studies have incorporated the use of CSR to help support both English-only and EL students. Two of these studies specifically included EL students (Klinger and Vaughn 2000; Klinger et al. 2004); however, only one analyzed specific outcomes for ELs (Klinger and Vaughn 2000). Looking specifically at how EL students helped each other while working in small, heterogeneous groups, Klinger and Vaughn (2000) provided 37 EL students in fifth grade with CSR intervention for 2–3 days a week for 4 weeks, for 30–40 min daily. These CSR lessons were specifically incorporated into science instruction. The researchers utilized a pre–posttest design to measure growth in vocabulary over time. Using two paired-sample *t*-tests, the authors reported statistically significant gains for students' overall vocabulary growth. No effect sizes were reported.

Clearly, both of these CSR studies with EL students demonstrate the potentially promising effects of using CSR to help improve ELs' reading and vocabulary. However, both of these studies have significant limitations in their design and execution that limit the generalizability of their findings. Both studies included small sample sizes, which resulted in low statistical power and affected the study's statistical conclusion validity (Shadish et al. 2002). Additionally, the Klinger and Vaughn (2000) study did not include a control group or random sampling or assignment of students to treatment, thereby affecting their ability to generalize the findings to other groups of students (Shadish et al. 2002). As such, more research should be conducted with EL students and CSR before recommending its widespread use.

Important features of tier 1 instruction include specific strategies to facilitate ELD and reading skill development. Among the most promising practices available for promoting ELD while facilitating access to academic content areas is the SIOP model. Additionally, one of the strongest evidence-based practices for facilitating reading skill development during core instruction is the use of differentiated instruction strategies. Two promising approaches for ELs include PALS and CSR. In the next section, effective screening practices are detailed for EL students.

Tier I Reading Screening for ELs

Assessment practices within an RTI model should be the same for ELs as are they are for monolingual students (Baker et al. 2010); in an

RTI model, assessments are typically used for screening, progress monitoring, and diagnostic purposes. Most often, curriculum-based measurements (CBMs) are used for screening and progress monitoring, and norm-referenced tests may be used for diagnostic decisions. In the specific case of ELs, regardless of a student's home language, children who are learning to read in English should be assessed in English to appropriately assess their acquisition of reading skills in their language of instruction (Kaminski and Good 2011). This is important because the purpose of the assessment is to predict the student's likelihood of learning to read in English, not learning to read in his or her primary language. As previously stated, it is critical to place ELs' reading performance within the context of their English language proficiency (August and Shanahan 2006), yet it is important to understand that, independent of their level of English language proficiency, screening measures are still good predictors of future reading performance (Lesaux and Siegel 2003). Based on this point, it is critical that ELs are included in reading screening as soon as possible, even if their English language proficiency has not reached a proficient level (Gersten et al. 2007).

While many have theorized that the same processes underlie the development of beginning reading skills across languages, as the *Test Standards* suggest, the use of specific tests for ELs should not be based on research conducted exclusively with NES (AERA/APA/NCME 1999). Recent studies have helped to determine the extent to which assessments developed with NES students similarly predicted reading outcomes for ELs. Lesaux and colleagues (Lesaux et al. 2007) conducted a longitudinal study of ELs representing 33 different languages who were assessed in kindergarten and fourth grade. In this study, the tests used in kindergarten included eight standardized and experimental measures of reading, namely the letter identification test, the Stanford-Binet working memory for sentences, sound mimicry, rhyme detection, syllable identification, phoneme deletion, oral cloze, and simple spelling (Lesaux et al. 2007). When the authors compared the growth of EL students' reading outcomes with NES' reading outcomes in fourth grade, they found that these measures, which are typically used to identify NES students who are at risk for reading concerns, were equally as predictive for ELs as they were for NES. Further, Leseaux and colleagues reported that "for both language groups, letter identification in kindergarten was predictive of initial mean differences and growth over time in word reading" (p. 831). Initially, EL students performed significantly worse on tests in kindergarten that were mediated by language skills (i.e., working memory, sound mimicry, rhyme detection, and oral cloze); however, by the fourth grade, both EL and NES performed similarly on tests which required well-developed English language skills (Lesaux et al. 2007).

While it is important to consider the results from Lesaux and colleagues (2007), the measures used in the study are measures that differ from those used in RTI models. Yet, recent work with CBM-type measures and ELs have shown similar results. In the area of early literacy, Vanderwood et al. (2008) conducted a longitudinal study examining the relationship between nonsense word fluency (NWF) scores from Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Kaminski and Good 1998) and performance on a statewide accountability measure of reading comprehension (i.e., California Achievement Test (CAT)). Across the sample of 134 ELs, the correlation between first-grade NWF scores and their scores in third grade on the CAT 6th Edition (CAT6) was $r=0.34$. This correlation was significantly higher than the relationship between a first-grade reading accountability measure (i.e., Stanford Achievement Test, 9th edition) and students' third-grade CAT6 scores ($r=0.17$). The correlation between the first-grade NWF score and the third-grade CAT6 scores was consistent across all levels of English fluency as determined by a statewide measure of English proficiency. These data suggested that, based on the results from this study, NWF, a commonly used screening measure in RTI, is an effective literacy screener for first-grade EL students. It is important to note that this study utilized data from 3 consecutive years, making this study unique with respect to methods and findings; there currently are no studies using data from English-only students that address these same research questions, so it is difficult to compare these findings.

Similar results have been found with measures of oral reading fluency (ORF). Early research by Baker and Good (1995) indicated ORF could be used to predict reading performance for Hispanic ELs and this finding has been systematically replicated. Several authors have provided evidence of the validity for using ORF with ELs to identify level of reading risk (Domínguez de Ramírez and Shapiro 2006; Wiley and Deno 2005). Most recently, Baker and colleagues (Baker et al. 2011) identified ORF could be used as a valid indicator of reading comprehension skills for second-grade ELs. Using confirmatory factor analysis, Baker et al. found a strong relationship ($r = 0.66$) between the Stanford Achievement Test (10th Ed.) and ORF for 96 students who attended schools participating in Reading First in the Pacific Northwest.

One area of concern related to the use of CBMs is the degree to which cut scores suggested by commercial publishers accurately reflect the probability of future success to the same degree for ELs as they do for NES. For example, if a specific score on a measure of ORF indicates a less than 80 % chance of a student reaching desired reading outcomes, it is not clear whether that same level of probability of success applies to ELs. There are some who argue cut scores should be developed specifically for ELs (Johnson et al. 2009; Vanderwood and Nam 2008), yet there is not sufficient research evidence to know how different the scores need to be based on English proficiency within EL groups. It is possible that one level of risk (and associated cut score) could be used for ELs who are substantially below desired levels of English proficiency, and another level of risk (and associated cut score) could be used for students who are more proficient in English. At this point, educators should be extra vigilant about using multiple sources of data to understand the degree to which an EL will or will not need supplemental support in reading.

Focusing on Tier 2: Intervention and Progress Monitoring for ELs

In recent years, there has been a proliferation of research on literacy interventions for ELs. After reviewing the available data, a fairly clear conclusion can be made: reading skills for Spanish-speaking ELs develop in a manner that is quite similar and consistent with how these skills develop for NES (Vaughn et al. 2005). In fact, several authors promote the use for ELs of the same five literacy components (i.e., phonemic awareness, alphabetic principal, fluency, vocabulary, and comprehension) as are promoted for all students by the National Reading Panel (NRP; Gersten et al. 2007; Vaughn et al. 2005; Vanderwood and Nam 2008). Again, it is very difficult to conclude whether the expectations about the effectiveness of the tools (i.e., interventions) can be applied with equivalent effects to all levels of English proficiency and all first languages (e.g., Spanish).

Early Literacy Interventions

As already mentioned, recently developed assessments can be used as early as the beginning of kindergarten to identify ELs who are at risk for reading difficulties that may be caused by underdeveloped phonological awareness skills and/or difficulty learning sound–symbol correspondences (Francis et al. 2006; Kaminski and Cummings 2007). These assessments are the same tools that are used with NES, yet the important question is whether similar intervention approaches can be used to remediate the skill deficits. Unfortunately, a significant amount of intervention research conducted before the past 5 years did not use rigorous designs that allow valid generalizations to be made. Yet, as the studies reviewed below indicate, initial evidence suggests early literacy skills can be remediated and result in performance trajectories for ELs that are similar to NES.

For example, in a randomized controlled trial designed to examine the impact of direct instruction reading interventions, Kamps and colleagues

(2007) determined ELs at risk for reading difficulties who were exposed to reading intervention made statistically significant gains in reading outcomes compared to ELs in a control group. The students in this study were enrolled in schools that were randomly assigned to a treatment or control condition. Out of 318 first- and second-grade children identified for their study, 148 were English-only students and 170 were EL students. The home language of the EL students was identified as primarily Spanish; however, some participants spoke Somalian, Sudanese, or Vietnamese.

In total, six schools participated, three in the treatment and three in the control condition. Students in the treatment and control schools were divided into at-risk and not-at-risk groups based on the results of three measures: DIBELS NWF, DIBELS ORF, and the Woodcock Reading Mastery Test (WRMT). Students in the treatment schools identified as at risk received tier 2 intervention in small groups (three to seven students) using direct instruction techniques with sequenced scripted lessons from four different reading curricula: Reading Mastery (SRA 1995), Early Interventions in Reading (Mathes and Torgesen 2005), Read Well (Sprick et al. 1998), and Read Naturally (Ihnot 2003). At-risk students in the other three schools (control condition) received tier 2 intervention using a balanced literacy approach that incorporated guided reading (e.g., Frey et al. 2005; Fountas and Pinnell 1996) and ELD in groups of 12 or more students (Kamps et al. 2007).

Results indicated that ELs in the three schools receiving the sequenced scripted lessons made significant gains in reading, and made greater gains than those students receiving the balanced literacy approach in the control schools on DIBELS measures (NWF $d=0.879$; ORF $d=0.94$) and the WRMT (all subtests $d>1.0$). Overall, Kamp and colleagues found that 50–60 % of students in the treatment condition were at or approaching benchmark on both NWF and ORF at the end of the intervention compared to 17 % of students in the control condition.

Gyovai et al. (2009) used a single-subject design to examine the effects of a supplemental early reading intervention on early literacy skills for 12 EL kindergarten students. The 12 students were identified as at risk on two indicators: the fall DIBELS benchmark and the Woodcock Johnson Tests of Achievement, Third Edition. The study utilized the Early Reading Intervention (Simmons and Kame'enui 2003), an intervention program designed to provide explicit instruction in phonics and phonological skills. A multiple-baseline across subjects design was implemented, and the four students with the lowest scores began to receive intervention for 20 min per day, 4 days per week. Once these students' scores began to improve (as measured by weekly progress monitoring through DIBELS), the next group of students began to receive intervention at the same level of support. Similarly, when these students' progress-monitoring scores demonstrated an increasing trend, the third group began to receive intervention two times per week. Intervention was delivered for 12–50 instructional sessions, depending on the student's group. The average level of change from baseline to the end of intervention was 17.3 sounds on Phoneme Segmentation Fluency subtest and 12.7 on the NWF subtest of DIBELS. The intervention produced effect sizes of more than 1.0 for 9 of the 12 students, indicating a large effect for the majority of participants (Gyovai et al. 2009).

Healy et al. (2005) also investigated the response of ELs to a supplemental, manualized phonological awareness curriculum using DIBELS. Fifteen first-grade students were identified as at risk based on fall screening with PSF and NWF. For the intervention, students were divided into small groups (five per group), an experimental phonological awareness intervention was provided two times per week for 30-min sessions, and progress was monitored with PSF and NWF each week. After 25 sessions, 12 of the 15 students met their PSF and NWF goals and were exited from the intervention; the 3 students who did not meet their goals were referred for more intensive tier 3 services. The average level of change for the 15 students was 38 sounds for PSF ($SD=13$) and 36.2 sounds for NWF ($SD=16.7$; Healy et al. 2005). It should be pointed out that this study is best perceived as a case study of the application of an RTI model with ELs due to the lack of use of an experimental, quasi-experimental, or single-case experimental design.

Fluency, Vocabulary, and Comprehension Interventions

Most intervention research with ELs has focused on providing phonics and/or phonological awareness interventions. However, some researchers have begun to examine the effect of fluency-building interventions on reading outcomes for ELs who are struggling readers. In their study of five Spanish-speaking, EL second-grade students, Ross and Begeny (2011) used a single-subject alternating treatment experimental design to examine the effects of a fluency intervention delivered at three levels: one-on-one individual instruction, small-group instruction, and a no-treatment control condition. Students were initially identified as at risk based on the DIBELS ORF benchmark assessment and the Test of Word Reading Efficiency (TOWRE; Torgesen et al. 1999), and received seven sessions of each intervention condition (one-on-one and small-group), and four sessions of the no-treatment control condition. Both the small-group and the one-on-one conditions included listening passage preview, repeated reading, retell, and phase-drill error correction. Multiple analyses, including a randomization test, indicated that both the small-group and one-on-one intervention improved students' scores on ORF. For one student, the randomization test indicated the one-on-one condition proved more successful in helping the student make significant gains, but for all of the others, there was no meaningful difference in students' reading outcomes between the small-group and one-on-one condition. Four of the five students produced ORF gains that were greater than average national growth norms and three students grew more than a half SD on the TOWRE. The researchers indicated that their findings in this study are similar to those of English-only students, and argued that fluency interventions may help ELs as much as English-only students (Ross and Begeny 2011). Due to the design of the study, it is challenging to determine whether the change in student scores were due exclusively to the intervention procedures or a combination of one-to-one and small-group interventions.

Progress Monitoring and ELs

As stated previously, the *Test Standards* (AERA/APA/NCME 1999) indicate that psychometric information specifically addressing the "linguistic" needs of the population being assessed should be evaluated to determine the appropriateness of using a measure with linguistically diverse students. There is evidence that most of the typical (e.g., CBM systems) screening and progress-monitoring tools can be given in a reliable manner to ELs (Baker and Good 1995; Healy et al. 2005), yet it is a challenge to conclude sufficient evidence exists to suggest that decisions made with progress-monitoring data are supported by enough evidence to be considered valid for ELs. One of the prerequisites to appropriately interpreting progress-monitoring data is to know the expected growth rate for students on the measure when adequate instruction is provided. Work conducted by Fuchs and colleagues (e.g., Fuchs et al. 1993) and others (AIMSweb 2012) provides growth rates on several progress-monitoring measures that can be used with NES, yet there is substantially less information about the appropriateness of these growth rates for ELs.

At this point, educators who use progress-monitoring tools to assess the impact of interventions with ELs should expect that for students who have a beginning level of English proficiency, growth rate on progress-monitoring measures will most likely be less than that of students with more developed levels of English language. However, there is reason to believe that with sufficiently intensive intervention, a group of ELs can grow at a faster rate than NES who are considered "at risk" readers (Lesaux and Siegel 2003). The challenge with making the conclusion about the appropriate rate of growth is due to the lack of documentation about the language proficiency of the EL sample in studies that focus on measuring growth with progress-monitoring tools (e.g., CBM; Vanderwood and Nam 2008).

Conclusion and Recommendations

The application of a multi-tier model with ELs is considered appropriate by several authors (e.g., Gersten et al. 2007), yet there are still many questions left unanswered. There is a clear need to develop diagnostic tools that will allow practitioners to identify the primary cause of a reading problem while taking into account a student's overall English language proficiency. At this stage, interpretation of any type of assessment result needs to be contextualized within a student's language background, including the extent to which the student has been formally (i.e., through instruction) and informally exposed to his or her home language and English. A system for conducting this contextualization needs to be developed and examined in rigorous studies that incorporate students who have varied language backgrounds so that targeted intervention planning can occur that is child and home language specific.

There is also a need to explore the utility of using RTI data to help support the special education eligibility process for specific learning disabilities (SLD). Although states are starting to implement RTI-focused eligibility models, at this point it is not clear how these models will affect SLD identification. It is quite possible that models that emphasize the use of response data will identify a different group of ELs for special education than current approaches, but at this stage this hypothesis must be viewed as an unanswered empirical question.

In addition to the two aforementioned areas that need more investigation, there is also a need to investigate whether the typical assessment and intervention tools used in RTI perform differently across differing levels of English proficiency. Most studies examining the application of the components of RTI with ELs are based on the results of samples that primarily consist of Spanish-speaking ELs. Another challenge with the current state of EL RTI research is that very few studies examined the application of the components for students learning to read languages other than English, with some rare and notable (but perhaps increasing) exceptions. For example, Vaughn and colleagues (2006) did demonstrate an ability to identify and intervene with at-risk students who were learning to read Spanish.

Therefore, the recommendations provided below are most well supported when the language of reading instruction is English and the first language of the student is Spanish. This is not to suggest the recommendations would not work in other situations or should not be attempted, but it seems inappropriate to attempt to say we know what will happen if these ideas are applied to all ELs. It is also important to note that the conclusions related to assessment are supported by a fair amount of research, but EL reading intervention research using strong empirical designs remains quite limited. With these limitations, the following recommendations for applying a multi-tier model with ELs are provided:

1. Provide educators a system for supporting the development of academic English during core instruction.
2. Research in this area is in its infancy, but educators may want to consider monitoring the development of academic English throughout the school year using targeted assessments created from the school's curricula.
3. When possible, provided differentiated ELD instruction that might include more intensive support to those students who are not developing English language skills at a desired rate.
4. Implement techniques for differentiating core reading instruction that place a focus on individualized and immediate instructional feedback (e.g., peer tutoring).
5. Screen all students on reading skills three times per year using typical RTI reading tools with norms or cut scores that account for differences in English language proficiency.
6. Routinely assess all ELs who receive reading intervention with typical RTI progress-monitoring tools and use the data to modify instruction.
7. When possible, provide reading intervention in homogenous groups based on English language proficiency.
8. Use the NRP framework of phonemic awareness, alphabetic principle, fluency, vocabulary,

and comprehension for conceptualizing and delivering reading interventions.
9. Examine the impact of all tiers by English proficiency at least once a year.
10. Use external resources (e.g., What Works Clearinghouse) and internal data when evaluating the effectiveness of reading interventions.

References

AERA, APA, & NCME (1999). Scales, norms, and score comparability. In AERA, APA, & NCME (Eds.), *Standards for educational and psychological testing* (pp. 49–60). Washington, DC: American Educational Research Association.

AIMSweb. (2012). Technical Manual. http://www.aimsweb.com/wp-content/uploads/aimsweb-Technical-Manual.pdf.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The condition of education 2011* (NCES 2011-033). *U.S. Department of Education, National Center for Education Statistics*. Washington, DC: U.S. Government Printing Office.

August, D., & Shanahan, T. (2006). *Executive summary: Developing literacy in second language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. New Jersey: Lawrence Erlbaum Associates.

Baker, S., & Good, R. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review, 24,* 561–579.

Baker, S., Simmons, D., & Kame'enui, E. (1997). Vocabulary acquisition: Research bases. In D. C. Simmons & E. J. Kame'enui (Eds.), *What reading research tells us about children with diverse learning needs: Bases and basics*. Mahwah: Erlbaum.

Baker, S. K., Gersten, R., & Linan-Thomson, S. (2010). Early reading instruction and intervention with ELLs: Key considerations in a multitiered approach. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (3rd ed., pp. 15–41). Bethesda: National Association of School Psychologists.

Baker, D. L., Stoolmiller, M., Good, R. H., III, & Baker, S. K. (2011). Effect of reading comprehension on passage fluency in Spanish and English for second grade English learners. *School Psychology Review, 40*(3), 331–351.

Calhoon, M. B., Al Otaiba, S., Cihak, D., King, A., & Avalos, A. (2007). Effects of a peer-mediated program on reading skill acquisition for two-way bilingual first-grade classrooms. *Learning Disability Quarterly, 30,* 169–184.

California Department of Education. (2012). *Facts about English language learners in California*. http://www.cde.ca.gov/ds/sd/cb/cefelfacts.asp. Accessed 26 Sept 2012.

Cummins, J. (1984). *Bilingual education and special education: Issues in assessment and pedagogy*. San Diego: College Hill.

Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research and Practice, 23*(2), 79–89.

Domínguez de Ramírez, R., & Shapiro, E. S. (2006). Curriculum-based measurement and the evaluation of reading skills of Spanish-speaking English language learners in bilingual education classrooms. *School Psychology Review, 35,* 356–369.

Echevevarria, J., Short, D., & Powers, K. (2006). School reform and standards-based education: A model for English-language learners. *The Journal of Educational Research, 99*(4), 195–210.

Echevarria, J., Vogt, M. E., & Short, D. J. (2008). *Making content comprehensible for English learners: The SIOP model*. New York: Pearson.

Echevarria, J., Vogt, M. E., & Short, D. J. (2010). *Making content comprehensible for secondary English language learners: The SIOP model*. Boston: Allyn & Bacon.

Fernald, A., & Weisleder, A. (2011). Early language experience is vital to developing fluency. In S. B. Newman & D. K. Dickenson (Eds.), *Handbook of early literacy research* (Vol. 3, pp. 3–19). New York: Guilford Press.

Fountas I. C., & Pinnell G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth: Heineman.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. (Under cooperative agreement grant S283B050034 for U.S. Department of Education). Portsmouth: RMC Research Corporation, Center on Instruction. http://www.centeroninstruction.org/files/ELL1-Interventions.pdf. Accessed 21 Sept 2012.

Frey, B. B., Lee, S. W., Tollefson, N., Pass, L., & Massengill, D. (2005). Balanced literacy in an urban school district. *The Journal of Educational Research, 98*(5), 272–280.

Fry, R. (2008). *The role of schools in the English language learner achievement gap*. Washington, DC: Pew Hispanic Center.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, C. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27.

Genesee, F., & Geva, E. (2006). Cross-linguistic relationships in working memory, phonological processes, and oral language. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the national literacy panel on language-*

minority children and youth. New Jersey: Lawrence Erlbaum.

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide* (NCEE 2007-4011). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee. Accessed 30 Sept 2012.

Gutierrez, G., & Vanderwood, M. L. (2013). A growth curve analysis of literacy performance among second-grade, Spanish-speaking, English-language learners. *School Psychology Review, 42*(1), 3–21.

Gyovai, L. K., Cartledge, G., Kourea, L., Yurik, A., & Gibson, L. (2009). Early reading intervention: Responding to the learning needs of young at-risk English language learners. *Learning Disability Quarterly, 32,* 143–162.

Hart, B., & Risley, R. T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.

Healy, K., Vanderwood, M., & Edelston, D. (2005). Early literacy interventions for English language learners: Support for an RTI model. *The California School Psychologist, 10,* 55–63.

Ihnot, C. (2003). *Read naturally*. St. Paul: Read Naturally.

Johnson, E., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24*(4), 174–185.

Jones, R. E., Yssel, N., & Grant, C. (2012). Reading instruction in tier 1: Bridging the gaps by nesting evidence-based intervention within differentiated instruction. *Psychology in the Schools, 49*(3), 210–218.

Kaminski, R., & Cummings, K. D. (2007). *DIBELS myths and facts*. *Dynamic Measurement Group*. http://dibels.org/papers/Myths_0208.pdf. Accessed 20 Oct 2012.

Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum based measurement* (pp. 113–142). New York: Guilford.

Kaminski, R., & Good, R. (2011). *Dynamic Indicators of Basic Early Literacy Skills: DIBELS Next Assessment Manual*. Eugene: Dynamic Measurement Group.

Kamps, D., Abbot, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., Culpepper, M., & Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier interventions. *Learning Disability Quarterly, 30,* 153–168.

Klinger, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal, 96,* 275–293.

Klinger, J. K., & Vaughn, S. (1999). Promoting reading comprehension, content learning, and English acquisition through collaborative strategic reading (CSR). *The Reading Teacher, 52*(7), 738–747.

Klinger, J. K., & Vaughn, S. (2000). The helping behaviors of fifth-graders while using collaborative strategic reading during ESL content classes. *TESOL Quarterly, 34,* 69–98.

Klinger, J. K., Vaughn, S., Arguelles, M. E., Hughes, M. T., & Leftwich, A. (2004). Collaborative strategic reading: “Real-world” lessons from classroom teachers. *Remedial and Special Education, 25*(5), 291–302.

Kosanovich, M., Ladinsky, K., Nelson, L., & Torgeson, J. (2005). Differentiated reading instruction: Small group alternative lesson structures for all students. *Florida Center for Reading Research*. http://www.fcrr.org/assessment/pdf/smallgroupalternativelesson-structures.pdf. Accessed 20 Oct 2012.

Lau v. N. (1974). 414 U.S. 563.

Leafstedt, J. M., & Gerber, M. M. (2005). Crossover of phonological processing skills: A study of Spanish-speaking students in two instructional settings. *Remedial and Special Education, 26*(4), 226–235.

Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39*(6), 1005–1019.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99*(4), 821–834. doi:10.1037/0022-0663.99.4.821.

Mathes, P., & Torgesen, J. (2005). *Early interventions in reading*. New York: McGraw Hill/SRA.

National Assessment on Educational Progress. (2009). Year in review: A compilation of results from the 2009 Mathematics, reading, Science, and High School Transcript Study reports. *National Center for Education Statistics Institute of Education Sciences*. http://nces.ed.gov/nationsreportcard/pdf/about/2011471.pdf. Accessed 20 Sept 2012.

National Clearinghouse for English Language Acquisition. (2011). *What language instruction educational programs do states use to serve English learners?* NCELA Fact Sheet. Washington, DC: Author. www.ncela.gwu.edu/files/uploads/NCELAFactsheets/refugees_2011.pdf. Accessed 2 Sept 2012.

Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students. (2012). *Biennial Report to Congress on the Implementation of the Title III State Formula Grant Program, School Years 2006–07 and 2007–08*. Washington, D.C.

Olsen, L. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for long-term English learners*. Long Beach: Californians Together.

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. New York: Guilford Press.

Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Terry, A. (2012). Past-tense morphology and phonological deficits in children with dyslexia and children with language impairment. *Learning Disability Quarterly, 46*(3), 230–240.

Ross, S. G., & Begeny, J. C. (2011). Improving Latino, English language learners' reading fluency: The effects of small-group and one-on-one intervention. *Psychology in the Schools, 48*(6), 604–618.

Saenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with disabilities. *Exceptional Children, 71*(3), 231–247.

Saunders, W., & Goldenberg, C. (2010). Research to guide English language development instruction. In F. Ong & V. Aguila (Eds.), *Improving education for English learners: Research-based approaches*. Sacramento: California Department of Education.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Short, D. J., Echevarria, J., & Richards-Tutor, C. (2011). Research on academic literacy development in sheltered instruction classrooms. *Language Teaching Research, 15*(3), 363–380.

Simmons, D. C., & Kame'enui, E. J. (2003). *Scott foresman early reading intervention*. Glenview: Scott Foresman.

Sprick, M., Howard, L., & Piddanque, A. (1998). *Read well*. Longmont: Sopris West.

Solari, E. J., & Gerber, M. M. (2008). Early comprehension instruction for Spanish-speaking English language learners: Teaching text-level reading skills while maintaining effects on word-level skills. *Learning Disabilities Research and Practice, 23*(4), 155–168.

The Nation's Report Card. (2011). *National Assessment of Educational Progress at Grades 4 and* 8. Institute of Education Sciences. U.S. Department of Education.

Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Center for Research on Education, Diversity & Excellence (CREDE). http://www.crede.ucsc.edu/research/llaa/1.1_final.html. Accessed 20 Sept 2012.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin: PRO-ED Publishing.

Vanderwood, M. L., & Nam, J. (2008). *Best practices in assessing and improving English language learners' literacy performance in best practices in school psychology V*. Maryland: NASP Publications.

Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review, 37*(1), 5–17.

Vaughn, S., Chard, D. J., Bryant, D. P., Coleman, M., Tyler, B., Linan-Thompson, S., & Kouzekanani, K. (2000). Fluency and comprehension interventions for third-grade students. *Remedial and Special Education, 21*(6), 325–335.

Vaughn, S., Klinger, J. K., & Bryant, D. P. (2001). Collaborative strategic reading as a means to enhance peer-mediated instruction for reading comprehension and content-area learning. *Remedial and Special Education, 22*(2), 66–74.

Vaughn, S., Mathes, P. G., Linan-Thompson, S., & Francis, D. J. (2005). Teaching English language learners at risk for reading disabilities to read: Putting research into practice. *Learning Disabilities Research & Practice, 20*(1), 58–67.

Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., & Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities, 39*(1), 56–73.

White, C., & Kim, J. (2009). *Putting the pieces together: How vocabulary instruction and expanded learning time can address the literacy gap*. Center for American Progress. http://www.americanprogress.org/wp-content/uploads/issues/2009/05/pdf/elt_may09.pdf. Accessed 20 Sept 2012.

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* 207–214.

Essential Features of Tier 2 and 3 School-Wide Positive Behavioral Supports

Barbara S. Mitchell, Allison L. Bruhn and Timothy J. Lewis

The social behavior challenges educators face on a daily basis across American schools has long been established. Overall levels of minor but chronic disrespect, noncompliance, and aggression continue to rise and are linked to later more intensive behavioral challenges (Benedict et al. 2007; Conroy et al. 2004; Duda et al. 2004; Muscott et al. 2009; Heaviside et al. 1998). Further, for those students who are identified as having an emotional/behavioral disorder (EBD) and receiving specialized instruction, intervention impact has been limited (Bradley et al. 2004; Wagner et al. 2005). Recent data indicate that over half of students receiving special education services under the EBD category fail to graduate with a high school diploma (Wagner et al. 2005), 20% have been arrested at least once while a student (VanAcker 2004), and the majority will require ongoing mental health and social assistance across their lifetimes (Walker et al. 2004).

To date, as described above, the overall impact of specialized supports for students with EBD have not been universally effective in altering significant behavioral challenges. However, services provided under the Individuals with Disabilities Education Act (IDEA) have lead to modest improvements (Wagner et al. 2005). A further challenge for educators tasked with providing specialized and intensive individualized supports is the overall gross under-identification within the IDEA category of "seriously emotionally disturbed" (EBD). Since passage of the law, fewer than 1 % of students have been identified as having EBD (U.S. Department of Education 2005). Experts within the EBD field estimate that 5–7 % of students manifest emotional and behavioral concerns significant enough to warrant special education services (Kauffman 2005; Walker et al. 2004). Simple mathematics indicates that using 5 % as the expected prevalence across the school-age population equals 2,810,149 children not receiving services under the Serious Emotional Disturbance (SED) label who might otherwise be eligible (U.S. Department of Education 2005). And while outcomes to date indicate improvements are needed in serving this population, the complete absence of specialized supports will surely exacerbate the social, emotional, and behavioral challenges among students at high risk.

The combination of increased overall levels of problem behavior and the large numbers

This chapter was supported in part by a grant from the Office of Special Education Programs, US Department of Education (H326S980003). Opinions expressed herein are those of the authors and do not necessarily reflect the position of the US Department of Education, and such endorsements should not be inferred. Inquires should be directed to Dr. Barbara Mitchell, 303 Townsend, University of Missouri, Columbia, MO 65211 (mitchellbs@missouri.edu).

B. S. Mitchell (✉) · T. J. Lewis
University of Missouri, Columbia, USA
e-mail: mitchellbs@missouri.edu

A. L. Bruhn
University of Iowa, Iowa City, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_31

of students at risk for significant emotional and behavioral challenges who currently may not be receiving specialized services has been a driving factor in establishing school-wide positive behavior support (SW-PBS) as both a preventative measure and universal intervention to support high-risk students (Sugai et al. 2000a). SW-PBS provides educators with a problem-solving framework that matches evidence-based practices to presenting problems unique to each school setting (see Chap. ## for a more in depth overview of SW-PBS). An additional feature of SW-PBS is the attention to creating school-wide systems that provide the necessary training and technical assistance to all faculty and staff within the school and ideally across the school district. School teams apply the problem-solving logic of SW-PBS: (a) data-based decisions to identify behavioral concerns and to progress monitor, (b) implementation of evidence-based practices matched to data, and (c) implementing systems of support to increase implementation fidelity across a continuum of available interventions creating a parallel to the academic response-to-intervention (RTI) multi-tiered system framework. Universal supports target common social behavioral challenges by explicitly teaching pro-social alternative behaviors, providing multiple opportunities for student practice, and providing high rates of positive specific feedback on skill use. In addition, universal expectations, instructional strategies, and environmental supports are adapted across structured (e.g., classroom) and unstructured (e.g., playground, cafeteria) school environments. Implementation of universal SW-PBS is monitored through student outcomes as well as annual fidelity checks (e.g., *School-wide Evaluation Tool,* Horner et al. 2004). For those students who are not successful with universal supports alone, a continuum of social, emotional, cognitive, and behavioral interventions are put in place through tier 2, or small-group interventions, and tier 3 individualized strategies.

The remainder of this chapter focuses on building a complete continuum of SW-PBS with emphasis on tiers 2 and 3. It is important to note that practices and systems of support overviewed within this chapter are intended to be part of an interconnected system of behavioral supports to create an integrated multi-tiered system of support (MTSS). Schools should be implementing universal supports with fidelity as a foundation to establishing a complete continuum of supports prior to systemic development of additional connected tiers. Ongoing attention to maintaining universal supports with high fidelity is also a critical prerequisite for successful installation of tier 2 and tier 3 interventions. An additional consideration in the establishment of tiers 2 and 3 is attention to the phases of implementation (see Chap. ##) and matching training and coaching for schools based on their readiness. Finally, within the continuum of SW-PBS, educators should view additional tiers beyond universal not as discrete and separate but rather as intensifying instruction that reinforces the universal support strategies (i.e., teaching, practicing, providing feedback, and modifying environments) provided at tier 1 to match the intensity of student need. Likewise, educators should not view tiers and students as synonymous. In other words, there are no "tier 2 students" but there will be students who will require additional levels of support across the school day to increase the likelihood they are successful.

Essential Features of Tier 2 and 3 SW-PBS Systems

Tier 2 and 3 supports within a multi-tiered system are designed to provide additional instruction and supports for students who are struggling to meet the goals and objectives taught in the core academic and behavioral curriculum (Chard 2013). Within SW-PBS, school teams are encouraged to implement a variety of interventions and supports based on student need. Regardless of the specific tier 2 and 3 supports used, as these interventions likely vary by school, there are several fundamental components associated with the identification of students requiring these levels of support and subsequent documentation of effectiveness. Essential features of tier 2 and 3 systems include: (a) insuring universal strategies are in place with high fidelity, (b) commitment from school administrators and a specialized behavior support team (tier 2/3 SW-PBS team) to lead intervention efforts, (c) a proactive method to identify in a timely manner the students who require additional supports, (d) a process for matching intervention with student need, (e) selection, adoption, and use of research-based interventions, (f) regular monitoring of student progress to assess performance and rate of improvement, and (g) coordinated decision-making to determine student movement (e.g., more or less intensive supports) within the multi-tiered system (Horner et al. 2010; Crone et al. 2010). Along with these key features, effective tier 2/3 systems also include plans for monitoring fidelity of implementation across interventions and of the overall system, as well as periodic evaluation of impact and effects relative to the goals of improving social and academic outcomes (see Fig. 1 for additional features).

Universal SW-PBS Implemented with Fidelity

Following the logic of RTI research within schools implementing SW-PBS, the process begins with all students having access to a rigorous and relevant academic and social–behavioral curriculum in an organized, effective, and positive classroom environment (Chard 2013; Lewis and Sugai 1999). In terms of behavior and social skills, all staff implement universal strategies that include teaching clearly defined expectations, rules, and procedures; explicit instruction for meeting school-wide behavioral goals; high rates of recognition for social and behavioral success; and consistent response for incorrect behavior that includes reteaching and practice opportunities for students who need these. These universal-level supports are implemented with fidelity by all staff, for all students, across all school settings to ensure each child participates in high-quality instruction before determining that he or she requires additional intervention. In particular, all classrooms consistently implement essential instructional management strategies, so that any additional supports or recommended environmental modifications as part of the tier 2/3 process will be positioned to have maximal impact (Simonsen et al. 2008). Specifically, in the same manner of having expectations and rules that are communicated consistently and taught school-wide across settings (e.g., cafeteria, hallway, restroom, playground), individual teachers also use an instructional approach for promoting desired behaviors and social skills within their classrooms. Productive classroom environments are established by using strategies that structure the learning environment so that a majority of problems are prevented from occurring. Strategies include using behavior-specific praise to recognize desired behaviors, providing frequent opportunities to respond during instruction, using material that is correctly matched with student's instructional level, providing choices to students (e.g., order of task completion), alternating easier tasks with those that are more challenging, pacing instruction adequately, implementing clear routines and procedures, and designing physical arrangements of the room that permit active supervision by adults (Kern and Clemens 2007; Simonsen et al. 2008).

Following adequate academic and behavioral instruction, assessment data are gathered and reviewed on a regular basis to evaluate each student's success in core instruction. Typically students are not considered for additional intervention (i.e., tier 2) until they have had sufficient time to respond to the universal strategies (e.g., approximately six to 8 weeks) implemented during core instruction. During this time, it is critical to confirm that universal interventions or core instructional strategies are implemented with fidelity.

Tier 2/3 SW-PBS Team

Initial development of SW-PBS begins with the formation of a team to focus on core instruction for social–behavioral skills. As the need for tier 2/3 intervention emerges, it is recommended that schools maintain the universal support-focused SW-PBS team but form a second group to address tiers 2 and 3. The tier 2/3 team should maintain

Fig. 1 Key features of effective Tier 2 systems of support within a continuum of SW-PBS. SW-PBS school-wide positive behavior support (Adapted from Crone et al. 2010)

Feature	*Description*
Continuous Availability	Tier2 interventions should be regularly available in the school such that students can be added to the intervention as soon as a need is identified, rather than waiting for supports to be developed.
Quickly and Easily Accessible	Ideally the Tier 2 is designed to allow rapid access to additional support. Optimally, implementation of the intervention occurs within 2-3 days of identification. This requires development of an efficient process for gathering and reviewing relevant student data that informs intervention selection.
Minimal Time Commitment Required From Classroom Teachers	All Tier 2 interventions require some level of commitment and implementation from classroom teachers. For example, teachers may be asked to modify traditional methods or implement new teaching practices such as increasing positive feedback, offering choices, or monitoring behavioral progress using a rating process. While teacher implementation is critical to student success ideally Tier 2 intervention components will fit within existing classroom routines, require minimal changes to existing routines, and call for only a few more minutes of teacher time each day.
Required Skill Sets Can Be Easily Learned	Strategies that require intensive training and skill development most often do not fall within the scope of Tier 2 interventions. Instead the Tier 2 techniques classroom teachers most commonly are asked to employ are consistent with quality instruction or are skills that can be easily learned. In fact, Tier 2 strategies frequently are an extension or higher intensity dosage of Tier I practices already in the repertoire of classroom teachers who implement SW-PBS.
Aligned With Universal School-Wide Strategies	Tier 2 interventions are based on and consistent with expectations and procedures already established within the universal school-wide system. Tier 2 programs typically offer a natural connect point with universal school-wide expectations by providing re-teaching along with additional prompts, cues, and practice opportunities for demonstrating expected behaviors. In addition, students can be recognized for appropriate behavior using the existing universal implementation plan the defining difference at Tier 2 is a higher rate of recognition.
All Personnel Are Aware of The Intervention(s) & Their Roles In The Process	All staff can name and describe the behavioral interventions available for students in their setting. In addition, staff members are aware of the process for seeking behavioral interventions for students about whom they have concerns. Staff members can explain their role and specific responsibilities associated with each Tier 2 behavioral intervention, can implement relevant components with fidelity, and receive regular feedback regarding their implementation and student response to the intervention. Staff members can interpret progress-monitoring data that contributes to data based decisions about interventions.
Consistent Implementation	A distinct advantage of the Tier 2 support process is the most common interventions can be implemented in the same manner across groups of students. This allows several students to have access to an evidence based intervention at the same time. Group-based delivery is an efficient way to meet the needs of many students who may be identified for behavioral risk. In addition, delivery of interventions in a similar manner across groups of students also may be associated with higher levels of implementation fidelity because participating staff members are asked to learn a specific set of skills per intervention rather than being required to deliver multiple individualized behavior plans.

linkages with the universal SW-PBS team by insuring all tier 2/3 strategies are couched within the set of school-wide expectations and that all staff are aware of their roles in supporting tier 2/3 interventions by including common members such as the building administrator, but also capitalize on existing behavioral expertise in the school, district, or regional SW-PBS initiative. Tier 2/3 teams often include school psychologists, special educators, behavior specialists, counselors, and one or more content-area specialists (e.g., language arts or mathematics).

Personnel who serve on the tier 2/3 teams establish systems and practices for students requiring more intensive social, emotional, and/or behavioral support. Members of this group ensure timely access to interventions, oversee implementation of selected interventions, regularly use data to monitor student progress during intervention, and evaluate overall program outcomes (Crone and Horner 2003). Membership of the team is determined according to ability and interest in fulfilling key roles and completing specific responsibilities. At a minimum, a tier 2/3 team typically includes one or more school administrators who are able to make decisions about staff, student, and school-wide schedules, reallocate resources, and set an overall tone of commitment and importance about the development, implementation, and ongoing monitoring of behavioral interventions. Also, it is recommended that the tier 2/3 team include staff with specific expertise in academic and behavioral assessment and intervention. While a primary focus of the team is development and implementation of social, emotional, and behavioral supports, many of the students identified for additional interventions will be at risk also for or already experiencing poor academic outcomes. The specialized behavior support team needs at least one member who can identify academic skill deficits and match students with appropriate academic intervention in addition to behavioral supports that may be needed. Finally, the tier 2/3 team should be structured to include general education teacher representation. Teachers are responsible often for making initial referrals for assistance, selecting appropriate strategies to meet student needs, and for implementing and monitoring effects of interventions (Debnam et al. 2013). Regardless of educational role, the more critical aspects when deciding on membership are structuring the team in a way that allows collaborative and data-based decision-making, division of a shared work load, and commitment to an instructional approach for behavior management and discipline (McIntosh et al. 2013)

Student Identification for Tier 2/3 Supports

Once SW-PBS practices, data-decision making and systems of support are established, implemented with fidelity, and have a measurable impact on student outcomes, tier 2/3 teams should develop a proactive process to actively seek out students who either are not responding to the core instruction or are displaying patterns of behavior that warrant additional intervention to lessen risk (e.g., Glover and Albers 2007; Kratochwill et al. 2004; Simmons et al. 2000; Walker and Shinn 2002). To identify students in need of tier 2/3 behavioral interventions, tier 2/3 teams develop a comprehensive system so that all students in a classroom, school, or district have an equal opportunity to be considered against characteristics of risk (e.g., acting out, argumentative, verbally or physically aggressive, overly shy, worried or withdrawn, anxious, unusually sad; see Fig. 2). The identification process is designed to promote early access to readily available interventions before student problems develop to a level that requires intensive intervention. It also includes clear criteria for indicating which students require immediate, intensive assistance (e.g., safety concerns for student or others). In addition, the identification approach is intended to identify students with internalizing (e.g., social withdrawal, anxiety) and externalizing (e.g., disruption, aggression) concerns regardless of the presence or absence of risk factors (Kamphaus et al. 2010).

Office Discipline Referral Data Within an SW-PBS framework, one commonly used measure of student social behavior performance is an office discipline referral (ODR), which is a standardized record of events of problem behavior. Documentation of disciplinary events is carried out consistently across students in that behavioral infractions are identified by common definitions and a standard set of information about these incidents is recorded and collected (i.e., type of problem, location, time of day, others involved, and possible motivation). Parallel to the tiered public health model that matches level of need with level of support (Gordon 1983), SW-PBS researchers and school teams frequently use number of ODRs accrued to define the level of support each student may require (McIntosh et al. 2009b). Regular review of ODR data is the

Fig. 2 Multi-method approach for early identification of social–emotional and behavioral risk

Method & Timeline	Process	Use of Results
School-based universal screening assessment using a research validated social-emotional behavior rating scale. • Annually – during school enrollment or registration process.	Parent or primary caregiver completes rating scale (e.g., Strengths & Difficulties Questionnaire [SDQ], Goodman, 1997)	Students with low to moderate risk scores are enrolled in a self-management intervention that includes communication between school and home (e.g., goal setting, behavioral contract, or a class-wide peer-assisted self-management program, Mitchem, Young, West, &Benyo, 2001) Students with moderate to high-risk scores are enrolled in behavioral intervention beginning the first week of school (e.g., First Step to Success, Walker et al., 1997; Check-in/Check-out, Crone, Hawken, & Horner, 2010; or Check & Connect).
School-based universal screening assessment using a research validated social-emotional behavior rating scale. • Biannually (e.g., early October and late January)	Classroom teacher completes rating scale for each student in class.	Students with scores in risk range who are not already receiving intervention are enrolled.
Review of existing school data (e.g., discipline, attendance, and health center/counselor visits) • Monthly	Data based criteria activates consideration for intervention (e.g., third discipline referral, absence or tardy during any grading period).	SWPBS Tier 2 team reviews relevant information. Students meeting one or more data based indicators who are not already receiving intervention are enrolled.
Teacher nomination • Continuously available	At any time school staff may nominate a student for social-emotional, behavioral support.	SWPBS Tier 2 team recommends one of the following: monitor progress, enroll in intervention, or nominate for intensive, individualized support.

primary method for determining impact of tier 1, core behavioral instruction. Review of ODRs also serves as a way to identify individual students who continue to display social behavioral problems even with exposure to high-quality core instruction (McIntosh et al. 2009). Considering that ODRs are already commonly collected, easily available and cost-efficient, use of these data for identifying students in need of tier 2 intervention may be more likely than use of other measures that may require additional funding or collection procedures (McIntosh et al. 2009). The following criteria, derived in part from proportions of ODR distributions in a large sample of schools, are currently viewed as a method for monitoring level of support required (Horner et al. 2005). Students receiving zero to one ODR per year are considered as responsive to tier 1. That is, the foundational level of support (i.e., structuring of environment to include instruction for behavioral expectations and recognition for successful demonstrations) provided to all students is sufficiently meeting the needs of students with low or no frequency of disciplinary events. Students with documented disciplinary events in the range of two to five incidents are considered for receiving tier 2 supports. Students with six or more ODRs are recommended for tier 3 interventions. Although ODRs offer valuable information about the frequency, topography, location, and potential motivation for behavior, they tend to be more representative of students with externalizing, rather than internalizing, behaviors (Walker et al. 2005).

Universal Screening Another method for systematically identifying students who may require additional support is use of a brief, behavioral screening instrument. Typically, screening instruments require a response to short statements about emotional or behavioral characteristics of a student. These instruments can be used to generate risk scores for all students in a classroom, grade level, building, or district.

There are several potential advantages for developing a systematic identification process that incorporates universal screening. First, screening instruments are generally perceived as a quick, accurate, and respectful process with capacity to include all children and youth of interest. Second, if an error occurs most often it is on the side of caution with the tendency to overidentify rather than missing students who are in need of support. Third, use of screening scores also informs schools about the needs of their particular student population which can assist with planning and resource mapping by finding groups of students with common needs. Finally, universal screening is recommended as an evidence-based practice by a number of different influential groups associated with educational policy and practice (e.g., President's Commission on Special Education 2002; No Child Left Behind Act 2001; U.S. Department of Health and Human Services, 2001).

Unfortunately, there are a number of reasons why universal screening for behavioral risk has not become a more common practice yet. The following list represents concerns that often are expressed (Levitt et al. 2007):

- Behavior is viewed as purposeful rather than as associated with environmental arrangements.
- Historically, schools tend to be reactive rather than proactive with respect to behavior.
- There is a widespread impression kids will "grow out of it" regarding problem behavior displayed during the early years of child development.
- Concerns about profiling or stigmatizing children and youth who meet risk criteria.
- Fear of costs and potential for identifying large numbers of students with EBD.
- General perception that it is easier to screen for vision and hearing concerns because the response falls in the realm of families.
- Political realities of managing parent reactions to behavior screenings and addressing issues of confidentiality.
- Lack of needed skill set. Educators often are not trained to respond to behavior with the same confidence that they are able to respond to academic concerns.

Within a tiered framework of support one important goal is to "catch" students before academic and/or behavioral challenges become severe. Universal screening provides an opportunity for all children to be considered for risk against identified criteria. This approach shifts focus from a traditional "wait to fail" service delivery model toward proactively seeking out children who may be at risk of academic failure and/or behavioral difficulties that would potentially benefit from specific instruction or intervention (Glover and Albers 2007). Use of a universal screening process also has the potential to minimize impact of risk and/or impede further development of more severe problems by identifying students before problems become severe.

The process for determining social behavioral risk within school settings is perhaps less firmly established than is the process for uncovering academic risk. Contemporary thinking in academic screening suggests that multiple concurrent methods are redundant and do not have a value-added effect. Instead, academic screening emphasizes gated or filtered screening with intervention trials in between. Yet related to behavioral concerns schools have long relied on teacher nominations (i.e., referral for special education) as the most common approach for identifying problems (Kamphaus et al. 2010). Increasingly, educators are also making use of commonly collected student behavioral data (i.e., ODR) (McIntosch et al. 2009; Walker et al. 2005). Systematic screening is becoming more prevalent but confronts educators with challenges of instrument selection, interpretation of results, capacity to meet the needs of identified students, and potentially negative perceptions of the process by families and community stakeholders who may view the procedures as invasive and/or stigmatizing (Golver and Albers 2007). *The President's Commission on Special Education* (United States Department of Education Office of Special Education and Rehabilitative Services 2002), the *No Child Left Behind Act of 2001* (NCLB; United

States Department of Education 2001), the US Public Health Service (2000), and the National Research Council (NRC; Donovan and Cross 2002) each indicate the need for early identification and intervention with recommendations to adopt universal, early, behavioral screening, yet limited information is available for schools regarding use of techniques and results (Albers et al. 2007)

The following section provides a brief description and sample items from several different screening questionnaires: (a) the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997), (b) the Behavioral and Emotional Screening System (BASC-2 BESS; Kamphaus and Reynolds 2007), and (c) the Systematic Screening for Behavior Disorders (SSBD; Walker and Severson 1992).

The Strengths and Difficulties Questionnaire

The SDQ is a brief behavioral screening questionnaire for children and youth aged 3–16 years old. All versions of the SDQ ask about 25 attributes, some stated positively and others negatively. These 25 items are divided between five scales: (a) emotional symptoms (five items), (b) conduct problems (five items), (c) hyperactivity/inattention (five items), (d) peer relationship problems (five items), and (e) prosocial behavior (five items). Scales (a) through (d) are added together to generate a total difficulties score (based on 20 items).

The same 25 items are included in questionnaires for completion by the parents or teachers of 4–16-year-old children (Goodman 1997). A modified informant-rated version is available for the parents or preschool teachers of 3- and 4-year-old children. In addition, questionnaires for self-completion by adolescents also are available and ask about the same 25 traits, with slightly different wording (Goodman et al. 1998). The self-report format is appropriate for youth aged 11–16. In general population samples, it is recommended to use a three-subscale division of the SDQ into internalizing problems, externalizing problems, and the prosocial scale (Goodman et al. 2010). The SDQ can be administered by hand and scored by hand or by entering scores on-line. Paper copies of the instrument can be downloaded and photocopies made with no charge. On-line administration and scoring for the SDQ also are available.

The Behavioral and Emotional Screening System (BASC-2 BESS)

The BASC-2 BESS offers a systematic way to determine the level of individual student risk drawn from ratings of behavioral and emotional strengths and weaknesses for considered students. The instrument was designed for children and adolescents in preschool through high school. The process consists of brief forms that can be completed by teachers, parents, or students individually or in any combination. Each rating form ranges from 25 to 30 items, requires no formal training for the raters, and is easy to complete, taking only 5–10 min of administration time per student. The screener assesses behaviors that represent both problems and strengths, including internalizing problems, externalizing problems, school problems, and adaptive skills. It yields one total score with associated risk classification (i.e., *normal, elevated, extremely elevated*) that is a reliable and accurate predictor of behavioral and emotional problems.

The Systematic Screening for Behavior Disorders

The Systematic Screening for Behavior Disorders (SSBD; Walker and Severson 1992) incorporates three gates, or stages. The screening takes into consideration both teacher judgments and direct observations in order to identify students at risk for developing ongoing internalizing and externalizing behavior concerns. Stage 1 of the SSBD involves teacher nomination. Stage 2 requires that teachers complete a critical events inventory and a short adaptive and maladaptive behavior checklist for each of the nominated students. Students whose scores on these checklists exceed the established cutoff are then candidates for stage 3. This final stage involves a 15-min interval observation in both the classroom and on the playground to determine a student's actual performance in social and classroom interactions.

Intervention Matched with Student Need

Once students are identified as nonresponsive to tier 1 instruction, the tier 2/3 team continues the SW-PBS logic by matching students with an appropriate level or tier of support. This decision is based on intensity, chronicity, and nature of the problem, as specific intervention strategies are selected to reflect individual student need (e.g., social skill deficit, self-management issue, emotional concerns). For example, a team may find a student has four ODRs and scored in the borderline range of total difficulties on the SDQ Goodman 1997), thus indicating the need for a tier 2 intervention. Or, a team may choose to forego tier 2 in favor of tier-3-level intervention for a student with more significant and intense problems (e.g., more than six ODRs, abnormal score on the SDQ). Beyond matching the intensity level (i.e., tier 2 or tier 3) of support based on student data, it is beneficial to consider particular attributes of the problem (e.g., function, location, setting, time of day, behavioral topography, acquisition deficit, performance deficit) prior to selecting an intervention (Hansen et al. 2014).

Accordingly, the tier 2/3 team should gather relevant information in a timely manner so that features of the problem are accurately identified, while still allowing for rapid access to interventions that are readily available. For a majority of identified students, screening results along with existing school data can be used to facilitate decisions that accurately inform intervention selection. Data that are easily accessible and generally useful for pinpointing aspects of social, emotional, or behavioral challenges may include documented disciplinary events (e.g., ODRs), student attendance patterns, grade point average and/or course grades, academic performance scores, and/or frequency of visits to the school nurse or counselor. These data can indicate when, where, and under what condition problem behavior is most likely to occur. For example, data may indicate a student is (a) reading below grade level, (b) often complains of sickness during reading class and thus frequently asks to see the nurse, and (c) when referred to the office for behavioral infractions the teacher consistently indicates "task avoidance" as a possible motivation for the problem behavior. Collectively, these data indicate the student may need both targeted small-group reading instruction, as well as behavioral supports to keep the student engaged during instruction. Similarly, universal screening tools originally used to identify students for tier 2/3 supports may provide relevant information for selecting an intervention. For instance, the SSBD may indicate a student has critical internalizing issues such as extreme anxiety. This information should be coupled with other data to determine the root of the anxiety, and in turn, select an intervention that confidently reflects the student's needs. When tier 2/3 teams triangulate data rather than relying solely on one tool or one statistic, they can more precisely capture the student's abilities and deficits, and thus match intervention accurately (Bruhn, Lane, & Hirsch, 2014; Marchant et al. 2009)

Beyond data collected as part of regular school practices, brief assessments such as the *Functional Assessment Checklist for Teachers & Staff* (FACTS; March et al. 2000) and the *Functional Assessment Screening Tool* (FAST; Iwata and DeLeon 1996) can be used to identify possible functions of students' problem behavior (e.g., access or avoidance of attention or tasks, sensory stimulation), which in turn, can inform the appropriate intervention selection. Although identifying function is generally regarded as part of the tier 3 process, quick tools such as the FACTS and FAST may help guide "function-based" thinking at tier 2. For example, if a student is misbehaving to gain peer attention, the student might be placed in a self-management intervention whereby if the student meets his/her goal of remaining on task they can access free time spent with peers (Briere and Simonsen 2011; Bruhn et al. 2014). Similarly, much of the research on the popular tier 2 intervention, check-in/check-out (CICO), has suggested students who are attention motivated are more likely to respond positively to CICO than students with escape-motivated behavior (Hawken et al. 2011; McIntosh et al. 2009).

For the few students (approximately 5% of the student population) who display intense and chronic problem behavior, a comprehensive functional behavioral assessment (FBA) should be conducted to create an equally individualized and intensive behavior support plan (BSP). Generally, students needing an FBA have long, complex histories of behavior problems and have been exposed to multiple risk factors (e.g., transiency, academic failure, substance abuse, poor parenting). Ideally, the team conducts interviews with the student's teachers, parents, and the student him/herself; reviews archival student records; and directly observes the student to identify antecedent conditions occasioning the target behavior and the consequences maintaining that behavior (Cooper et al. 2007). Data gleaned from the FBA are then used to design a highly individualized BSP. Tier 3 interventions such as function-based BSP's tend to require more time, effort, and resources than tier 2 interventions because they are acutely tailored to the individual rather than addressing a small group of students with comparable problems through more generalized supports as in tier 2. Function-based interventions have a record of success that began in clinical settings with students who had low-incidence disabilities, but more recently, positive outcomes have been documented across a range of school-based settings for students with persistent behavioral problems (McIntosh et al. 2008; Lane et al. 2009).

Regardless of the interventions the team has to choose from, teams must have adequate time to consider and plan for all students identified as needing support beyond tier 1. Use of a specific format for collecting, reviewing, and discussing relevant student information can keep conversations directed toward existing school-based supports and/or accessible community agencies and programs while avoiding discussion of factors that are beyond influence of the support team (e.g., homelife, community circumstances, previous experiences with related families). Maintaining a problem-solving approach that is focused on alterable indicators of risk through the use of research-based interventions will maximize time allotted for the behavior support team to address all students identified for additional support.

Research-Based Interventions

Following student identification and data review to determine appropriate intervention and level of environmental supports, intervention implementation, and ongoing supports should be provided. The tier 2/3 team should first determine their capacity to provide a range of interventions and supports, start with those currently in place, and expand to create a full range within each tier of the continuum. Key factors related to intervention implementation within a continuum of supports include: (a) clear alignment with core behavioral instruction and procedures (e.g., addresses school-wide behavioral expectations, uses same or similar language for recognizing or correcting behavior, follows school-wide reinforcement system, but provides a greater frequency of support), (b) a designated coordinator for each tier 2 strategy and designated case managers for students receiving tier 3 supports, (c) ongoing data review to assess implementation fidelity and student progress, and (d) a plan for fading or intensifying supports. Within tier 2, the current research base advocates (a) additional small-group social skill instruction addressing skill and performance deficits (Elliott and Gresham 2008), (b) empirically validated self-management strategies that have demonstrated improvements in social and academic behaviors such as CICO (Crone et al. 2010), Check and Connect (Christenson et al. 2012), or Check, Connect, and Expect (Cheney et al. 2009), and (c) academic supports either through differentiated and supplemental instruction ideally within an MTSS with progress monitoring (Epstein et al. 2008).

Research studies reporting positive impact of tier 2 interventions are evident. For example, outcomes from use of social skill instructional groups have shown decreases in disruptive behavior, increases in on-task behavior, improved scores from teacher ratings of social behavior and academic competence, along with increases in prosocial play skills, peer interactions, and communication (Kamps et al. 2011; Gresham et al. 2006; Marchant et al. 2007). Evidence for the group-oriented self-management program show similar results. For example, investigations of the CICO intervention consistently demonstrate decreases in ODR for student participants (e.g., Hawken et al. 2007; Todd et al. 2008), increases in academic engagement (Hawken and Horner 2003; Campbell and Anderson 2011), and reduced frequency of disruptions or negative social interactions (e.g., Campbell and Anderson 2008; McIntosh et al. 2009). Finally, a recent study that documented use of an academic support intervention that showed improvements in "assignment attack" behaviors such as task persistence and organization (Ness et al. 2011).

Tier 3 supports are guided by an FBA whereby an individual support plan is developed to teach a functionally equivalent replacement behavior and environmental modifications are made to reinforce replacement skill use (Gage et al. 2012). Additional mental health, family, and other non-educational supports are also included in tier 3 supports when indicated (e.g., RENEW; Malloy et al. 2010).

In addition to building small-group strategies within tier 2 supports, the tier 2/3 team can also include simple alterations within classroom and other school environments as an additional tier 2 option. It may be the case that a small-group intervention is not necessary to address the presenting concerns, rather intensifying universal practices in response to the problem may be sufficient to produce behavior change. For example, increasing universal strategies on targeted behaviors such as (a) providing more frequent prompts for expected behavior, (b) increasing the rate of specific performance feedback for correct behavior, (c) reteaching classroom rules and routines, or (d) altering the environment to increase supervision or the add effective instructional practices may be sufficient to meet the student's need.

Monitoring Student Progress

In the same manner that curriculum-based measures are one type of progress monitoring tool to assess academic performance over time, behavioral assessment data within SW-PBS can also be collected and used as the basis for determining students' RTI. In academic progress monitoring, there are explicit decision rules to guide service delivery and determine student responsiveness (VanDerHeyden et al. 2007). Behavioral progress monitoring, on the other hand, lacks an equivalent standard protocol. Rather, instructional programming decisions may be made based on data collected via systematic direct observation, direct behavior ratings (DBR; Chafouleas 2011), and review of archival data (e.g., documented disciplinary events, time out of instruction due to problem behavior, academic work samples). These are common techniques for monitoring student progress before, during, and after intervention.

Direct observation is certainly the most time and labor intensive progress monitoring option, as it requires an individual to watch a student's behavior for a set period of time and record either the numerical (e.g., frequency) or temporal dimension (e.g., duration) of the behavior. This process must be done frequently and data should be graphed for analysis. Additionally, to establish reliability of direct observation data, another observer may simultaneously, but independently, collect data using the same preestablished procedures. Then, the data may be compared for interobserver agreement (IOA). Although direct observation of both problem and appropriate behavior would provide schools with the best data source to inform instructional decisions, the challenges of (a) observing behavior expected across multiple educational settings, (b) the costs in terms of time to complete, and (c) the expertise required to collect reliable data often limit direct observation as a viable option for progress monitoring. This may be true especially for teachers who are asked to simultaneously teach and collect data, as many teachers view this as an impossible task (Epstein 2010). Instead, teacher perceptions as measured through informal and formal behavioral rating scales, and DBRs are more common approaches to monitoring behavioral interventions. A DBR is similar to direct observation in that it requires the teacher to directly observe the student. However, observation is not continuous and is accumulated over a set period of time (e.g., first period, circle time, 45 min). The DBR can be used in the pre-populated form, which asks teachers to rate students on a scale of 0–10 on three behaviors—academic engagement, disruptive, and respectful. Or, the teacher can tailor the DBR to reflect the behavior(s) of interest. Researchers have suggested the pre-populated DBR is moderately to highly correlated with frequency and duration data obtained via direct observation (Chafouleas et al. 2009; Riley-Tillman et al. 2008). Thus, DBR may be a more practical option than direct observation for teachers.

Data-Based Decisions for Movement Between Tiers

Whether the team selects direct observation or a DBR, regular review of student data is key to making decisions about movement within the tiered model. For example, in some cases, review of student data will demonstrate a student has consistently met a specific behavioral performance target over a period of several weeks. In this event, the Tier 2 team would have sufficient evidence to advocate for a gradual reduction in intervention supports that leads the student to a self-management and maintenance phase (e.g., Campbell and Anderson 2011). Or, regular review of student data might reveal highly variable performance or performance that is consistently below the expected level, representing a questionable RTI. Under these circumstances, the team will first verify fidelity of intervention implementation and then make decisions that could include simple modifications to the existing support or recommendation for a higher intensity intervention (e.g., Campbell and Anderson 2008; Fairbanks et al. 2007). Finally, in some cases, review of student performance data will indicate an overall poor response to the intervention that is demonstrated by an increase in frequency or intensity of problems and/or failure to improved at the expected rate or to the desired level of performance after sufficient exposure to an intervention matched with need and implemented with integrity. This type of data can serve as the basis for recommending additional information gathering (i.e., FBA) and development of an individualized support plan in tier 3 (Crone and Horner 2003; March and Horner 2002).

Monitoring Fidelity

With respect to fidelity of treatment, the tier 2/3 team has two foci. First, the tier 2/3 team must routinely consider how well or to what extent interventions are being accurately implemented. Second, the team should conduct, at least annually, a review of their SW-PBS tier 2/3 system that includes evaluation of overall effects and impact toward improvement of social and academic outcomes.

Tier 2/3 Intervention Fidelity Only when an intervention is implemented as designed can conclusions about behavior changes be made accurately (Gansle and Noell 2007; Yeaton and Sechrest 1981). That is, school teams that measure implementation quality and accuracy can have greater confidence in decisions made when reviewing student data. For example, if a student demonstrates limited or poor response during an intervention but the school team has measured and determined fidelity of implementation to be high, then it is likely the student truly is in need of adapted, alternate, or more intensive support. Conversely, if a student did not have the opportunity to benefit from intervention because it was not implemented with fidelity, then the team may need to provide additional training and resources to the interventionist and allow time for the intervention to be implemented with fidelity prior to placing the student in a more intense level of support (Bruhn et al. 2014; Mitchell et al. 2011). Additionally, teams must realize there are multiple factors such as intervention complexity, training adequacy, and skill of the interventionist that affect intervention integrity (Yeaton and Sechrest 1981). Regardless of the reason for low treatment integrity, performance feedback is critical for improvement and accurate implementation (Duhon et al. 2009; Keller-Margulis 2012). Performance feedback should involve the team reviewing fidelity data and participating in a constructive dialogue about setting goals for improving implementation and specific steps to take toward meeting those goals (Keller-Margulis 2012). Determining the extent to which all parts of an intervention are implemented accurately is a key priority for the tier 2/3 team. SW-PBS tier 2/3 teams are challenged with developing a process that (a) provides adequate evidence the intervention was implemented as intended, and (b) can feasibly be conducted on a regular basis. Three common methods for measuring accuracy of intervention implementation include: (a) intervention-specific product review (e.g., monitoring forms, rating scales, daily progress report (DPR), self-management records), (b) direct observation of intervention implementation, and (c) self-report measures completed by intervention providers. Each method includes both benefits and potential limitations.

Intervention Product Review One way to verify delivery of intervention is by completing a review of products associated with the intervention (Crone et al. 2010). For example, many tier 2 behavioral interventions include daily or weekly documentation of student performance such as a DPR system, as is used in CICO. In these cases, the behavior support team can examine three to five of the most recent progress documents to determine whether some of the specific elements of the intervention occurred (e.g., progress was recorded, progress was calculated and evaluated using specified criteria, data were shared with relevant stakeholders such as classroom teacher and family as evidenced by stakeholder signatures). When review of student products provides evidence that specified components of the intervention are in place and being delivered consistently, the team can have greater confidence in the accuracy of progress monitoring data and subsequent decisions made from use of those data. Alternately, when review of intervention-related products identifies an area of low implementation, a member of the support team can be designated to provide coaching, remediation of skills, and feedback about intervention delivery as needed (e.g., student, teacher, and/or parent). However, one limitation of the product review is that it does not take into account components of the intervention that are not part of the product, or in this case, the DPR. For instance, when students and interventionists use the DPR, it is expected that students will receive verbal feedback from an adult about their performance as well as a predetermined reward for meeting goals. These components, which may be critical to the success of the intervention, may need to be documented via direct observation or self-report (Bruhn, McDaniel, & Kreigh, in press.)

Direct Observation A second method for verifying accuracy of intervention integrity is direct observation. In these cases an observation checklist may be especially useful both for documenting features that occurred and for providing feedback to implementers. Direct observation tools can be developed to reflect the essential features of intervention delivery. For example, a tier 2 social skills intervention observation checklist might include: (a) instructor introduced, defined, and discussed skill use and its importance, (b) instructor demonstrated at least two examples and nonexamples of the skill, (c) students correctly demonstrated skill use in prompted role plays, (d) students correctly demonstrated skill use in nonprompted role plays, and (e) instructor reinforced occurrences of the appropriate skill and corrected incorrect skill use (Elliott and Gresham 2008). As a second example, within a tier 3 FBA-BSP intervention, direct observation targets might include: (a) teacher prompted student regarding correct skill use, (b) teacher consistently reinforced desired behavior, and (c) teacher minimized reinforcement of problem behavior. Like direct observation used to monitor student progress, direct observation of fidelity is also time and labor intensive, but is the most accurate method for estimating intervention integrity.

Self-report Measures A final option for collecting fidelity of implementation data is asking intervention providers to record components they provide and/or self-assess their level of accuracy according to a list of described features. Self-report measures can be organized to collect implementer ratings (e.g., five-point scale) or as a simple “yes” or “no” checklist. Interviews and questionnaires may be used to give informants an opportunity to elaborate beyond the scope of what is covered in a checklist or rating scale. Although self-report measures are often used to assess implementation because they are easy and require little time to complete, teams using this technique should be cautious when reviewing and evaluating data. Some research on self-reporting has indicated implementers tend to overrate their performance and accuracy (Wickstrom et al. 1998). However, more recently, some researchers have demonstrated self-report of fidelity can be an accurate and more efficient alternative to direct observation (Hagermoser Sanetti and Kratochwill 2009). Regardless, a multi-method, multi-informant approach to measuring fidelity that includes both direct and indirect observation can only increase the accuracy of conclusions drawn about intervention effects (Lane et al. 2004).

Tier 2/3 Systems Fidelity A hallmark of the SW-PBS process is coordinated universal, tier 2, and tier 3 systems of support (Lewis et al. 2010). Similar to the work of VanDerHeyden et al. (2007) who evaluated fidelity of an entire RTI system including not only intervention fidelity but also fidelity to the decision-making process; recent advancements in SW-PBS related to effective tier 2/3 intervention efforts led to the development of instruments designed to assess the fidelity of the overall systems, or process, of implementation. These tools are used both as metrics of fidelity as well as data sources for making data-based decisions to continually build and refine tier 2/3 systems. One example, the *Benchmarks for Advanced Tiers* (BAT; Anderson et al. 2010) was created to answer three main questions: (1) Are the foundational (organizational) elements in place for implementing tier 2 and 3 behavior support practices? (2) Are the essential features of a tier 2 support system in place? and (3) Are the essential features of a tier 3 system in place? The BAT is completed by the tier 2/3 team and reflects the consensus or majority of team member perceptions. A second example, the *Individual Student Systems Evaluation Tool* (ISSET; Anderson et al. 2012) is also used to measure the implementation status of tier 2/3 systems within a school (e.g., Debnam et al. 2013). The ISSET consists of 35 items and is divided into three parts: foundations, tier 2 interventions, and tier 3 interventions. A summary score is obtained for each of the three parts and is administered/completed by an external evaluator. Two data sources are used to score the ISSET: Administer and teacher interviews and a review of permanent products/documented procedures. Results from either instrument are used to identify areas of strength and needed improvements.

Evaluating Impact

Tier 2/3 teams, in concert with the universal SW-PBS, adhere to the basic logic of SW-PBS (i.e., data–practices–systems) and also conduct informal evaluations of the effects of their efforts. Teams focus on three questions: (a) Did the interventions we put in place lead to improved student outcomes? (b) If interventions were not effective, what else do we need to know to increase the likelihood of success (e.g., what training and technical assistance do we need)? (c) Could we have worked more efficiently (see Fig. 3 for a more comprehensive list of system evaluation points). To answer these questions, the tier 2/3 and universal SW-PBS teams determine what change has occurred across the variables or behaviors of interest. In the case of behavioral intervention, programs were likely selected with the expectation of impacting problem behavior and student engagement, which in turn may lead to improvements in academic achievement. Evaluating features of tier 2 and 3 program impact can occur using a variety of methods. For example, in some cases compilation of existing school data provides valuable evidence of program impact and can be used to evaluate system outcomes.

Under other circumstances, a more formalized assessment of behavior support team member perceptions or measures of system-wide implementation may be warranted when student outcomes are inconsistent (e.g., *Self-assessment Survey, SAS; School Safety Survey, SSS; Team Implementation Checklist,* TIC). School staff members typically complete the SAS survey at minimum annually. After reviewing outcomes from tier 2 or 3 interventions that demonstrated

Fig. 3 Sample evaluation questions to conduct tier 2/3 system review

- How many students participated in the tier 2/3 supports during the school year?
- How many of the participants were successful subsequently "graduated" from the intervention or support and what data indicated success?
- How many participants required adaptations to existing interventions?
- Were there students who required additional and/or more intensive supports? If so, how many students and what types of supports?
- From all students who at some point during the school year qualified to participate in a Tier 2 or 3 behavioral intervention how many also were at some point evaluated for special education eligibility?
- Were any specific subgroups of children served at disproportionate rates in tier 2/3 supports (e.g. culturally, linguistically, socio-economically, or ethnically diverse populations)?
- Were outcomes similar across all student groups?

questionable positive impact for multiple students, the team could review SAS survey results to get a better understanding about staff perceptions of implementation at the school-wide, nonclassroom, classroom, and individual student levels (Safran 2006). SAS results also demonstrate staff perceptions of priorities for improvement. Regular review of these data lends itself to ensuring the SW-PBS efforts across all tiers address the needs identified by staff that completed the survey.

In addition to the outcomes derived from existing school data, teams also gather social validation data. Social validity data typically provides a picture of the extent to which particular stakeholder groups (i.e., students, families, and teachers) value identified practices and outcomes. Social validity data are commonly gathered through use of a survey or asking personnel to respond to items on a brief questionnaire. Example statements or questions may include: *Overall problem behaviors have decreased for this student during participation in the behavioral intervention program; I think this behavioral intervention program may be good for other students in our school; Having my child/student in the behavioral intervention program is worth my time and effort.* If social validity results are low it may be difficult to continue implementation of the support "as is." Instead, teams will investigate why the practice is perceived poorly and make adjustments either by providing additional information and technical assistance and/or by making changes to features that perhaps were not feasible to maintain (Fixsen et al. 2009; Guskey 2002).

Finally, timelines for conducting the system evaluation must be considered also. An annual review that occurs near or after the end of each school year may be practical and make sense for many school teams. This time frame allows widespread participation in the system across many staff members, students, and parents throughout the school year, concludes during a period when student data are already commonly collected, and guides decisions the team will make for refining the system the following year. Annual evaluation allows time before the start of the next school year for making adjustments to the existing system such as improving the communication system to all staff regarding implementation if student outcomes are inconsistent across the school day or additional training for intervention implementers if fidelity is low. Systems evaluation requires thoughtful but realistic consideration. The process should use existing data collection strategies (e.g., student outcomes) and fidelity/planning tools (e.g., the BAT) to insure teams engage in the process. At the same time, evaluations also should not be so simplistic that valuable outcomes are overlooked or never uncovered.

Empirical Support for Tier 2/3 SW-PBS

To date, there is empirical evidence on the social, emotional, behavioral, and academic impact of tier 1 SW-PBS supports (Barrett et al. 2008; Bradshaw et al. 2010; Bradshaw et al. 2008; Curtis et al. 2010; Horner et al. 2009; Simonsen et al. 2010). Likewise, there is a strong evidence base demonstrating positive effects from use of particular interventions that are commonly implemented within tier 2 and 3 levels of support. For example, social skill instruction, self-management strategies, academic supports, and use of FBA data to guide individual BSPs have been verified as effective treatments for changing behavior (e.g., Bessette and Wills 2007; Christensen et al. 2007; Skinner et al. 2009; Sumi et al. 2013). However, empirical demonstrations of the impact of tier 2 and 3 supports within the context of a complete continuum of SW-PBS to date are limited. The following provides a brief summary of what has been concluded from the existing body of work.

Two recent reviews of tier 2 implementation studies examined questions associated with identification of students, delivery of supports, and overall impact (Bruhn et al. 2014; Mitchell et al. 2011). From the two reviews several themes emerged. First, on balance, the overall number of studies investigating tier 2 level of implementation has increased recently. Ranging from approximately 2002 through 2012 researchers conducted and published nearly 30 separate investigations concerning tier 2 interventions and delivery within a tiered framework (Bruhn et al. 2014). Second, demonstrations reporting sufficient fidelity of tier 1 implementation prior to initiation of tier 2 intervention, particularly at the classroom level, were fewer than expected (Bruhn et al. 2014). Third, although there is evidence of reliance on a few commonly collected data such as disciplinary events, screening scores, and teacher nominations as primary sources for determining which students will access intervention, wide variation in identification approaches were still clearly evident (Bruhn et al. 2014; Mitchell et al. 2011). Fourth, across the existing body of work, evidence generally indicated high social validity ratings for the variety of interventions that have been included. The remainder of this section provides further detail for some of these findings.

Core Instruction Delivered With Fidelity Within the SW-PBS literature, several instruments have been developed and used for monitoring implementation efforts. Examples include the TIC (Sugai et al. 2009), the SSS (Sprague et al. 2002), and the *Benchmarks of Quality* (BoQ; Cohen et la. 2007). Two of the most commonly used tools are the *School-Wide Evaluation Tool* (SET; Horner et al. 2004) and the *Effective Behavior Support/Self Assessment Survey* (EBS/SAS; Sugai et al. 2000). The SET is designed to assess features of tier 1 that are in place and evaluate ongoing efforts for developing a school-wide approach for behavior management and discipline.

The SET consists of interviews, observation of the setting, and a review of products such as teaching plans, signage, and data collection. An overall score is calculated and results of 80 % or higher indicate adequate implementation. In some of the initial publications for tier 2/3 interventions, it was common to simply state that features of tier 1, core instruction such as clearly defined expectations and rules, explicit behavioral instruction across settings, use of a recognition system, and consistent documentation of behavioral infractions were in place (e.g., Lane et al. 2002; Lane et al. 2003; Gresham et al. 2006). However, more recent investigations reflect the importance of high-quality core instruction as part of the continuum of supports by attending to and reporting scores from the SET (e.g., Campbell and Anderson 2011; Hawken et al. 2011; Mong et al. 2011; Simonsen et al. 2011). While the SET is a research-validated instrument with strong psychometric properties demonstrating consistency in measurement of tier 1 features, one aspect of tier 1 that is not addressed within the SET is how well the core behavioral instruction is provided within individual classrooms (Horner et al. 2004). Likewise, acknowledgment of students demonstrating expected behaviors and matching instruction with student ability are associated with sustained implementation and positive student outcomes (e.g., ODR) (Matthews et al. 2014). However, the existing literature related to tier 2 has not explicitly reported that scores from the SAS showed integrity of classroom-level implementation prior to delivery of behavioral supports for identified students (Bruhn et al. 2014; Mitchell et al. 2011).

Identification Practices and Tools The existing evidence also indicates some commonalities among tools that are used to identify students for tier 2 behavioral interventions, but still shows wide variation in decision criteria. For example, Mitchell et al. (2011) reported participation in a tier 2 intervention was based on one or a combination of the following: (a) a nomination process in which a classroom teacher, a parent, or a problem-solving team identified the student as at-risk; (b) use of existing behavioral performance data—typically, ODR information—to indicate that the student was unresponsive to the tier 1 prevention efforts or continuing to demonstrate difficulties meeting social behavioral expectations; or (c) use of a behavioral screening score. Similarly, Bruhn et al. (2014) determined (a) ODR; (b) teacher nomination; (c) academic performance; or (d) other methods such as parent nomination, attendance data, or a specified behavioral function served as tools for identifying student candidates for intervention. Yet, specific details regarding exact criteria differed across studies. Descriptions of several examples follow.

Nomination

Several studies asked teachers to identify, without a quantifiable index, which students needed intervention (Bruhn et al. 2014). In one study, participants were identified for intervention based on teacher nominations because of behavioral difficulties in the classroom and/or existence of a behavior plan (McCurdy 2007). A different example indicated teachers nominated students because of classroom problem behavior and lack of responsiveness to the tier 1 prevention efforts (McIntosh 2009). In a third example, a multi-informant process was used, which included administrator nomination, teacher verification of problem behavior, parental consent, and student willingness to participate, to determine which children would receive intervention (Todd et.al. 2008).

Student Data

The existing research base also provides examples for using specified cut points of student academic and/or behavioral data that are commonly collected in school settings to identify intervention candidates. For example, students who met a predetermined threshold of risk such as five or more disciplinary incidents or two or more documented events of problem behavior were selected for intervention (Hawken 2006; Hawken et al. 2007). Other studies have also included use of academic data such as grade point averages (e.g., Robertson and Lane 2007) or Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores (e.g., Wills et al. 2010) along with the collected behavioral data (Bruhn et al. 2014).

Behavioral Screening

In the most recent comprehensive review of tier 2 within SW-PBS, use of one or more screening instruments to identify students at risk was evident in less than half of all the reviewed studies (i.e., 13 out of 28, Bruhn et al. 2014). When a systematic screening process was used, the most commonly employed instruments were the Student Risk Screening Scale (SRSS; Drummond 1994), the SSBD (Walker and Severson 1992), and the Social Skills Rating System (SSRS; Gresham and Elliott 1990). Several examples combined use of the SSBD with the SRSS (e.g., Lane et al. 2008; Lane et al. 2010; Little et al. 2010) which is interesting considering the SSBD has been shown to identify both externalizing and internalizing concerns and at this time the SRSS is best known for detecting externalizing attributes alone. One study demonstrated use of a behavioral screener, the SSBD, but also included DIBELS as part of the identification method (Wills et al. 2010). Demonstration of screening as a method for tier 2 identification appears to be increasing within recent years. Among the 13 studies from the Bruhn et al. (2014) review, all studies that included a behavioral screening process were published within the past 10 years (i.e., 2003–2011) and a majority occurred in studies published in 2007 or later. While use of screening to identify students at risk in academic domains is well established, teams working within SW-PBS have continued to rely on more of a multi-method approach that often includes use of a psychometrically sound rating instrument, but may review other data such as informal teacher or school personnel perceptions of problems and/or recorded disciplinary events (Bruhn et al. 2014)

Social Validity

A final topic that was evident within both of the recently published tier 2 literature reviews was the issue of social validity. In a 1976 address presented for the Division of Experimental Analysis of Behavior within the American Psychological Association, Montrose Wolf described, with humor, how Don Baer helped prioritize "social importance" as it relates to the mission of applied behavior analysis and how subsequently ABA came to "have a heart" (Wolf 1978). Not quite 40 years later, the measurement of behavioral change that is perceived as socially valued or important is more common than not at least in the area of SW-PBS research. Currently the majority of studies conducted with tier 2 level interventions have reported social validity data (Bruhn et al. 2014; Mitchell et al. 2011). Most of the studies that included social validity measurements at minimum gathered these data from participating teachers who perhaps most closely detected the impact of school-based interventions (e.g., Fairbanks et al. 2007; Hawken et al. 2007). Several studies also demonstrated measurement of student beliefs about interventions that were delivered (e.g., Lane et al. 2002; Lane et al. 2003). Commonly used tools included the *Intervention Rating Profile* (IRP; Witt and Elliott 1985), the *Children's Intervention Rating Profile* (CIRP; Witt and Elliott 1985), and the CICO Program Acceptability Questionnaire (Crone et al. 2010). Results from these assessments across a variety of stakeholders showed generally favorable outcomes for indicators like ease of implementation, perceived benefit, and potential sustainability of practices over time (Bruhn et al. 2014).

Recommendations for Future Research

There is agreement about the types of practices to include in a full continuum of behavioral supports and there is understanding of the essential system features that will facilitate high-quality implementation. What is needed is a more expansive body of empirical evidence indicating positive outcomes (e.g., student academic and behavioral indicators, organizational health, school safety and climate) are attained and can be sustained over time when interconnected tiered support is delivered. In addition, research is needed on several factors within the tier 2/3 system including (a) impact on students with internalizing problem behavior (e.g., depression, anxiety), (b) the ability of teams to effectively and efficiently match intervention to presenting problem with existing commonly collected school data, (c) the impact of academic and behavioral supports guided by student need, (d) essential implementation steps that lead to sustained outcomes, and (e) the mediating and/or moderating effect of related environmental variables such as school district/state support, community and cultural factors, and prior student learning history. This will require sophisticated investigations demonstrating a clear linkage between tier 1 systems, including classroom, implemented with high-fidelity and student-need-directed tier 2/3 supports. While the emerging literature on impact of tier 2/3 supports within the context of SW-PBS is encouraging, the true "value add" has yet to be empirically demonstrated. The logic of nesting tier 2/3 supports within a connected and related instructional environment is built on decades of research on promoting maintenance and generalization of intervention outcomes and to identify and support students at the first sign of risk. Two recent studies provide a starting point to empirically validate the impact of both the prevention logic of SW-PBS and the impact of a complete continuum of supports.

First, Reddy and colleagues completed a meta-analytic review of the types of prevention and intervention programs available for students with or at risk for emotional and behavioral disorders (Reddy et al. 2009). Findings indicated that prevention and early intervention programs can be effective for reducing risk of onset and for minimizing impact of already existing symptoms. Second, Nelson et al. (2009) assessed child outcomes when a full continuum of behavioral supports was provided. The tiered system of supports consisted of Behavior and Academic Support and Enhancement (BASE) at the universal level (Nelson et al. 2002), First Step to Success as the tier 2 level intervention (Walker et al.1997), and multi-systemic therapy (MST) as the tier 3 treatment (Henggeler et al. 1998) and were tested with 407 children in grades K–3. Effects on problem behavior, social skills, and academic competence across levels of intervention (tier 1, tier 2, and tier 3) were measured. One finding of particular importance was that children who participated in the tier 2 or tier 3 interventions early in their school career showed immediate social behavioral improvements that could be sustained over time with continued implementation of the tier 1 supports (Nelson et al. 2009). However, the results did not demonstrate a similar outcome for teacher ratings of academic competence among students who received the advanced tier interventions. The results offered initial evidence to support the idea that a tiered service delivery model could impact behavioral performance of children with or at risk for behavioral problems (Nelson et al. 2009).

Implications for Practice

Across this chapter, essential features and examples to build SW-PBS systems at the tier 2/3 level of support have been provided. The limitations of a single chapter and the complexities of both specific intervention strategies and the intensity of support across school environments preclude simple recommendations for practice. The encouraging news as previously discussed, intervention and data collection strategies along with the necessary systems of support at the universal level are well established and comprehensive implementation guides are readily available for practitioners (see pbis.org). In addition, many districts have adopted integrated frameworks to tie in community supports at the individual student level. Unfortunately, schools and districts continue to struggle with establishing effective tier 2 systems of support. While the research base is limited with respect to systems, as noted above there is strong empirical evidence for specific intervention strategies appropriate for tier 2 level of supports (e.g., CICO, social skill instruction).

References

Albers, C. A., Glover, T. A., & Kratochwill, T. R. (2007). Introduction to the special issue: How can universal screening enhance educational and mental health outcomes? *Journal of School Psychology, 45*(2), 113–116.

Anderson, C., Lewis-Palmer, T., Todd, A., Horner, R., Sugai, G., & Sampson, N. (2008). Individual student systems evaluation tool. Unpublished instrument, *Eugene, OR: Educational and Community Supports, University of Oregon*. http://www.pbis.org/common/cms/files/pbisresources/ISSET_TOOL_v_3_March_2012.pdf. Accessed 1 March 2015.

Anderson, C. M., Childs, K., Kincaid, D., Horner, R. H., George, H., Todd, A. W., Sampson, N. K., & Spaulding, S. (2010). Benchmarks for advanced tiers (BAT). Unpublished instrument. Eugene OR: Educational and Community Supports. http://www.pbis.org/common/cms/files/pbisresources/BAT_v2.5.pdf. Accessed 1 March 2015.

Anderson, C. M., Lewis-Palmer, T., Todd, A. W., Horner, R. H., Sugai, G., & Sampson, N. K. (2012). *Individual student systems evaluation tool*. Educational and Community Supports, University of Oregon.

Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland statewide PBIS initiative systems, evaluation, and next steps. *Journal of Positive Behavior Interventions, 10*(2), 105–114.

Benedict, E., Horner, R. H., & Squires, J. (2007). Assessment and implementation of positive behavior support in preschools. *Topics in Early Childhood Special Education, 27*(3), 174–192.

Bessette, K. K., & Wills, H. P. (2007). An example of an elementary school paraprofessional-implemented functional analysis and intervention. *Behavioral Disorders, 32*(3), 192–210.

Bradley, R., Henderson, K., & Monfore, D. A. (2004). A national perspective on children with emotional disorders. *Behavioral Disorders, 29,* 211–223.

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of school-wide Positive Behavioral Interventions and Supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31*(1), 1–26.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12*(3), 133–148.

Briere, D. E. III, & Simonsen, B. (2011). Self-monitoring interventions for at-risk middle school students: The importance of considering function. *Behavioral Disorders, 36,* 129–140.

Bruhn, A. L., McDaniel, S., & Kreigh, C. (in press). Self-monitoring interventions for students with behavior problems: A review of current research. *Behavioral Disorders.*

Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2014). A review of tier 2 interventions conducted within multitiered models of behavioral prevention. *Journal of Emotional and Behavioral Disorders, 22*(3), 171–189.

Campbell, A., & Anderson, C. M. (2008). Enhancing effects of check-in/check-out with function-based support. *Behavioral Disorders, 33*(4), 233–245.

Campbell, A., & Anderson, C. M. (2011). Check-in check-out: a systematic evaluation and component analysis. *Journal of applied behavior analysis, 44*(2), 315–326.

Chafouleas, S. M. (2011). Direct behavior rating: A review of the issues and research in its development. *Education and Treatment of Children, 34,* 575–591. doi:10.1353/etc.2011.0034.

Chafouleas, S. M., Kilgus, S. P., & Hernandez, P. (2009). Using direct behavior rating (DBR) to screen for school social risk: A preliminary comparison of methods in a kindergarten sample. *Assessment for Effective Intervention, 34,* 224–230. doi:10.1177/1534508409333547.

Chard, D. (2013). Systems impact: Issues and trends in improving school outcomes for all learners through multitier instructional models. *Interventions in School and Clinic, 48*(4), 198–202.

Cheney, D. A., Stage, S. A., Hawken, L. S., Lynass, L., Mielenz, C., & Waugh, M. (2009). A 2-year outcome study of the check, connect, and expect intervention for students at risk for severe behavior problems. *Journal of Emotional and Behavioral Disorders, 17*(4), 226–243.

Christensen, L., Young, K. R., & Marchant, M. (2007). Behavioral intervention planning: Increasing appropriate behavior of a socially withdrawn student. *Education and Treatment of Children*, *30*(4), 81–103.

Christenson, S. L., Stout, K., & Pohl, A. (2012). *Check & connect: A comprehensive student engagement intervention*. Minneapolis: University of Minnesota.

Cohen, R., Kincaid, D., & Childs, K. E. (2007). Measuring school-wide positive behavior support implementation development and validation of the benchmarks of quality. *Journal of Positive Behavior Interventions, 9*(4), 203–213.

Conroy, M. A., Hendrickson, J. M., & Hester, P. (2004). Early identification and prevention of emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 199–215). New York: Guildford Press.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). New Jersey: Pearson.

Crone, D. A., & Horner, R. H. (2003). *Building positive behavior support systems in schools: Functional behavioral assessment*. New York: Guilford Press.

Crone, D., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program* (2nd ed.). New York: Guilford Press.

Curtis, R., Van Horne, J. W., Robertson, P., & Karvonen, M. (2010). Outcomes of a school-wide positive behavioral support program. *Professional School Counseling, 13*(3), 159–164.

Debnam, K. J., Pas, E. T., & Bradshaw, C. P. (2013). Factors influencing staff perceptions of administrator support for tier 2 and 3 interventions: A Multilevel perspective. *Journal of Emotional and Behavioral Disorders, 21*(2), 116–126.

Donovan, M. S., & Cross, C. T. (2002). Minority students in special and gifted education. Washington, DC: National Academy Press.

Dowdy, E., Doane, K., Eklund, K., & Dever, B. V. (2013). A Comparison of teacher nomination and screening to identify behavioral and emotional risk within a sample of underrepresented students. *Journal of Emotional and Behavioral Disorders, 21*(2), 127–137.

Drummond, T. (1993). *The student risk screening scale (SRSS)*. Grants Pass: Josephine County Mental Health Program.

Duda, M. A., Dunlap, G., Fox, L., Lentini, R., & Clarke, S. (2004). An experimental evaluation of positive behavior support in a community preschool program. *Topics in Early Childhood Special Education, 24*(3), 143–155.

Duhon, G. J., Mesner, E. M., Gregerson, L., & Witt, J. C. (2009). Effects of public feedback during response to intervention team meetings on teacher implementation integrity and student academic performance. *Journal of School Psychology, 47,* 19–37.

Elliott, S. N., & Gresham, F. M. (2008). *SSIS intervention guide*. Minneapolis: NC Pearson.

Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, R. (2008). *Reducing behavior problems in the elementary school classroom: A practice guide* (NCEE #2008–012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides. Accessed 23 Jan 2013.

Everett, S., Sugai, G., Fallon, L., Simonsen, B., & O'Keeffe, B. (2011). *School-wide tier 2 interventions: Check-in/check-out getting started workbook*. Center on Positive Behavioral Interventions and Supports Center for Behavioral Education and Research University of Connecticut.

Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M. (2007). Response to intervention: Examining classroom behavior support in second grade. *Exceptional Children, 73*(3), 288–310.

Fantuzzo, J. W., Rohrbeck, C. A., & Azar, S. T. (1987). A component analysis of behavioral self-management interventions with elementary school students. *Child & Family Behavior Therapy, 9,* 33–43.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*(5), 531–540.

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. *Behavioral Disorders, 37*(2), 55–77.

Gansle, K. A., & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VenDerHeyden. (Eds.) *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 244–251). New York: Springer.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38,* 581–586.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry, 7,* 125–130.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology, 38,* 1179–1191.

Gordon, R. S. (1983). An operational classification of disease prevention. *Public Health Reports, 98,* 107–109.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines: AGS Publishing.

Gresham, F. M., Van, M. B., & Cook, C. R. (2006). Social skills training for teaching replacement behaviors: Remediating acquisition deficits in at-risk students. *Behavioral Disorders, 31*(4), 363–377.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3), 381–391.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Treatment integrity assessment in the schools: An evaluation of the treatment integrity planning protocol. *School Psychology Quarterly, 24*(1), 24–35.

Hansen, B. D., Wills, H. P., Kamps, D. M., & Greenwood, C. R. (2014). The effects of function-based self-management interventions on student behavior. *Journal of Emotional and Behavioral Disorders, 22*(3), 149–159.

Hawken, L. S. (2006). School psychologists as leaders in the implementation of a targeted intervention: The Behavior Education Program. *School Psychology Quarterly, 21*(1), 91.

Hawken, L. S., & Horner, R. H. (2003). Evaluation of a targeted intervention within a schoolwide system of behavior support. *Journal of Behavioral Education, 12*(3), 225–240.

Hawken, L. S., & Hess, R. S. (2006). School psychologists as leaders in the implementation of a targeted intervention: The behavior education program. *School Psychology Quarterly, 21*(1), 91.

Hawken, L. S., MacLeod, K. S., & Rawlings, L. (2007). Effects of the behavior education program (BEP) on office discipline referrals of elementary school students. *Journal of Positive Behavior Interventions, 9*(2), 94–101.

Hawken, L. S., O'Neil, R. E., & McLeod, K. S. (2011). An investigation of the impact of function of problem behavior on effectiveness of the behavior education program (BEP). *Education and Treatment of Children, 34,* 551–574. doi:.org/10.1353/etc.2011.0031.

Heaviside, S., Rowand, C., Williams, C., & Farris, E. (1998). *Violence and discipline problems in U.S. Public Schools: 1996–97* (NCES 98-030). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Henderson, J., & Strain, P. S. (2009). *Screening for delays and problem behavior. (Roadmap to effective intervention practices)*. Tampa: University of South Florida.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford Press.

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions, 6*(1), 3–12.

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2005). *School-wide positive behavior support. Individualized supports for students with problem behaviors: Designing positive behavior plans* (pp. 359–390). New York: Guilford Press.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*(3), 133–144.

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42*(8), 2–14.

Iwata, B. A., & DeLeon, I. G. (1996). *The functional analysis screening tool (FAST)*. Gainesville: University of Florida, The Florida Center on Self-Injury.

Kamphaus, R. W., & Reynolds, C. R. (2007). *Behavioral and emotional screening system. Manual*. Minneapolis: Pearson.

Kamphaus, R. W., DiStefano, C., Dowdy, E., Eklund, K., & Dunn, A. R. (2010). Determining the presence of a problem: Comparing two approaches for detecting youth behavioral risk. *School Psychology Review, 39*(3), 395–407.

Kamps, D., Wills, H. P., Heitzman-Powell, L., Laylin, J., Szoke, C., Petrillo, T., & Culey, A. (2011). Class-wide function-related intervention teams: Effects of group contingency programs in urban classrooms. *Journal of Positive Behavior Interventions, 13*(3), 154–167.

Kauffman, J. M. (2005). How we prevent the prevention of emotional and behavioural difficulties in education. In P. Clough, P. Garner, J. T. Pardeck, & F. K. O. Yuen (Eds.), *Handbook of emotional and behavioural difficulties* (pp. 429–440). London: Sage.

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools, 49,* 342–352.

Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools, 44*(1), 65–75.

Kratochwill, T. R., Albers, C. A., & Steele Shernoff, E. (2004). School-based interventions. *Child and Adolescent Psychiatric Clinics of North America, 13*(4), 885–903.

Lane, K. L., & Mariah Menzies, H. (2010). Reading and writing interventions for students with and at risk for emotional and behavioral disorders: An introduction. *Behavioral Disorders, 35*(2), 82.

Lane, K. L., Wehby, J. H., Menzies, H. M., Gregg, R. M., Doukas, G. L., & Munton, S. M. (2002). Early literacy instruction for first-grade students at risk for antisocial behavior. *Education and Treatment of Children, 25*(4), 438–458.

Lane, K. L., Wehby, J. H., Menzies, H. M., Doukas, G. L., Munton, S. M., & Gregg, R. M. (2003). Social skills instruction for students at risk for antisocial behavior: The effects of small-group instruction. *Behavioral Disorders, 28*(3), 229–248.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential-but often forgotten-component of school-based interventions. *Preventing School Failure, 48,* 36–43.

Little, M. A., Lynne Lane, K., Harris, K. R., Graham, S., Story, M., & Sandmel, K. (2010). Self-regulated strategies development for persuasive writing in tandem with schoolwide positive behavioral support: Effects for second-grade students with behavioral and writing difficulties. *Behavioral Disorders, 35*(2), 157.

Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*(2), 163–191.

Lane, K. L., Kalberg, J. R., Bruhn, A. L., Mahoney, M. E., & Driscoll, S. A. (2008). Primary prevention programs at the elementary level: Issues of treatment integrity, systematic screening, and reinforcement. *Education and Treatment of Children, 31*(4), 465–494.

Lane, K. L., Bruhn, A. L., Crnobori, M., & Sewell, A. L. (2009). Designing functional assessment-based interventions using a systematic approach: A promising practice for supporting challenging behavior. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Policy and practice: Advances in learning and behavioral disabilities* (Vol. 22). Bingley: Emerald.

Lewis, T. J., Jones, S. E., Horner, R. H., & Sugai, G. (2010). School-wide positive behavior support and students with emotional/behavioral disorders: Implications for prevention, identification and intervention. *Exceptionality, 18*(2), 82–93.

Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive schoolwide management. *Focus on Exceptional Children, 31*(6), 1–24.

Liaupsin, C. J., Umbreit, J., Ferro, J. B., Urso, A., & Upreti, G. (2006). Improving academic engagement through systematic, function-based intervention. *Education and Treatment of Children, 29*(4), 573–591.

McCurdy, B. L., Kunsch, C., & Reibstein, S. (2007). Secondary prevention in the urban school: Implementing the Behavior Education Program. *Preventing School Failure: Alternative Education for Children and Youth, 51*(3), 12–19.

McIntosh, K., Horner, R. H., Chard, D. J., Dickey, C. R., & Braun, D. H. (2008). Reading skills and function of problem behavior in typical school settings. *Journal of Special Education, 42,* 131–147.

McIntosh, K., Campbell, A. L., Carter, D. R., & Dickey, C. R. (2009). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions, 11*(2), 82–93.

McIntosh, K., Reinke, W. M., & Herman, K. E. (2009a). School-wide analysis of data for social behavior problems: Assessing outcomes, selecting targets for intervention, and identifying need for support. In G. G. Peacock, R. A. Ervin, E. J. Daly, & K. W. Merrell (Eds.), *The practical handbook of school psychology* (pp. 135–156). New York: Guilford.

McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009b). Concurrent validity of office discipline referrals and cut points used in schoolwide positive behavior support. *Behavioral Disorders, 34*(2), 100–113.

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing research-based trajectories of office discipline referrals for individual students. *School Psychology Review, 39*(3), 380.

McIntosh, K., Mercer, S. H., Hume, A. E., Frank, J. L., Turri, M. G., & Mathews, S. (2013). Factors related to sustained implementation of schoolwide positive behavior support. *Exceptional Children, 79*(3), 293–311.

Malloy, J. M., Sundar, V., Hagner, D., Pierias, L., & Viet, T. (2010). The efficacy of the RENEW model: Individualized school-to-career services for youth at risk of school dropout. *Journal of at Risk Issues, 15*(2), 17–25.

March, R. E., & Horner, R. H. (2002). Feasibility and contributions of functional behavioral assessment in schools. *Journal of Emotional and Behavioral Disorders, 10*(3), 158–170.

March, R. E., Horner, R. H., Lewis-Palmer, T., Brown, D., Crone, D., Todd, A. W., & Carr, E. (2000). *Functional Assessment Checklist for Teachers and Staff (FACTS)*. Eugene: Educational and Community Supports.

Marchant, M. R., Solano, B. R., Fisher, A. K., Caldarella, P., Young, K. R., & Renshaw, T. L. (2007). Modifying socially withdrawn behavior: A playground intervention for students with internalizing behaviors. *Psychology in the Schools, 44*(8), 779–794.

Mathews, S., McIntosh, K., Frank, J. L., & May, S. L. (2014). Critical features predicting sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions, 16*(3), 168–178. 1098300713484065.

Mitchell, B. S., Stormont, M., & Gage, N. A. (2011). Tier two interventions implemented within the context of a tiered prevention framework. *Behavioral Disorders, 36*(4), 241–261.

Mong, M. D., Johnson, K. N., & Mong, K. W. (2011). Effects of check-In/check-out on behavioral indices and mathematics generalization. *Behavioral Disorders, 36*(4), 225–240.

Muscott, H. S., Pomerleau, T., & Szczesiul, S. (2009). Large-scale implementation of program-wide positive behavioral interventions and supports in early childhood education programs in New Hampshire. *NHSA Dialog, 12*(2), 148–169.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning: The effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders, 10*(3), 136–148.

Nelson, J. R., Hurley, K. D., Synhorst, L., Epstein, M. H., Stage, S., & Buckley, J. (2009). The child outcomes of a behavior model. *Exceptional Children, 76*(1), 7–30.

Ness, B. M., Sohlberg, M. M., & Albin, R. W. (2011). Evaluation of a second-tier classroom-based assignment completion strategy for middle school students in a resource context. *Remedial and Special Education, 32*(5), 406–416.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

Presidents Commission on Excellence in Special Education. (2002). A new era: Revitalizing special education for children and their families. Washington, DC: US Department of Education. www.ed.gov/inits/commissionsboards/whspecialeducation/index.html. Accessed 7 March 2013.

Reddy, L. A., Newman, E., De Thomas, C. A., & Chun, V. (2009). Effectiveness of school-based prevention and intervention programs for children and adolescents with emotional disturbance: A meta-analysis. *Journal of School Psychology, 47*(2), 77–99.

Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A. M., & Glazer, A. D. (2008). Examining the agreement of Direct Behavior Ratings and Systematic Direct Observation for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10,* 136–143. doi:10.1177/1098300707312542.

Robertson, E. J., & Lane, K. L. (2007). Supporting middle school students with academic and behavioral concerns: A methodological illustration for conducting secondary interventions within three-tiered models of support. *Behavioral Disorders, 33*(1), 5–22.

Safran, S. P. (2006). Using the effective behavior supports survey to guide development of schoolwide positive behavior support. *Journal of Positive Behavior Interventions, 8*(1), 3–9.

Simmons, D. C., Kuykendall, K., King, K., Cornachione, C., & Kameenui, E. J. (2000). Implementation of a school-wide reading improvement model: "No one ever told us it would be this hard!". *Learning Disabilities Research & Practice, 15*(2), 92–100.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children, 31*(3), 351–380.

Simonsen, B., Britton, L., & Young, D. (2010). School-wide positive behavior support in an alternative school setting: A case study. *Journal of Positive Behavior Interventions, 12*(3), 180–191.

Simonsen, B., Myers, D., & Briere, D. E. (2011). Comparing a behavioral check-in/check-out (CICO) intervention to standard practice in an urban middle school setting using an experimental group design. *Journal of Positive Behavior Interventions, 13*(1), 31–48.

Skinner, J. N., Veerkamp, M. B., Kamps, D. M., & Andra, P. R. (2009). Teacher and peer participation in functional analysis and intervention for a first grade student with attention deficit hyperactivity disorder. *Education and Treatment of Children, 32*(2), 243–266.

Sprague, J., Colvin, G., & Irvin, L. (2002). *The school safety survey. The Institute on Violence and Destructive Behavior*. University of Oregon College of Education.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R., Wickham, D., Reuf, M., & Wilcox, B. (2000a). Applying positive behavioral support and functional behavioral assessment in schools. *Journal of Positive Behavioral Interventions, 2,* 131–143.

Sugai, G., Horner, R. H., & Todd, A. W. (2000b). *Effective behavior support: Self-assessment survey*. Eugene: University of Oregon.

Sugai, G., Horner, R. H., & Lewis-Palmer, T. (2009). The team implementation checklist (v. 3.0). *Unpublished instrument. Eugene, Oregon: Educational and Community Supports, University of Oregon*.

Sumi, W. C., Woodbridge, M. W., Javitz, H. S., Thornton, S. P., Wagner, M., Rouspil, K., Yu, J. W., Seeley, J. R., Walker, H. M., Golly, A., Small, J. W., Feil, E. G., & Severson, H. H. (2013). Assessing the effectiveness of first step to success: Are short-term results the first step to long-term behavioral improvements? *Journal of Emotional and Behavioral Disorders*, *21*(1), 66–78.

Todd, A. W., Campbell, A. L., Meyer, G. G., & Horner, R. H. (2008). The effects of a targeted intervention to reduce problem behaviors: Elementary school implementation of check in—check out. *Journal of Positive Behavior Interventions*, *10*(1), 46–55.

United States Department of Education Office of Special Education and Rehabilitative Services. (2002). *A new era: Revitalizing special education for children and their families*. Washington, DC: United States Department of Education Office of Special Education and Rehabilitative Services.

United States Public Health Service. (2000). *Report of the surgeon general's conference on children's mental health: A national action agenda*. Washington, DC: Department of Health and Human Services.

U.S. Department of Education. (2005). 27th annual report to Congress on the implementation of the Individuals with Disabilities Education Act, 2004. Washington, DC: Author.

US Department of Health and Human Services. (2001). Report of the Surgeon General's conference children's mental health: A national action agenda. Washington, DC: Author.

VanAcker, R. (2004). Current status of public education and likely future directions for students with emotional and behavioral disorders. In L. M. Bullock & R. A. Gable (Eds.), *Quality personnel preparation in emotional/behavioral disorders: Current perspectives and future directions*. Denton: Institute for Behavioral and Learning Differences.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders (SSBD): User's guide and administration manual*. Longmont: Sopris West.

Walker, H. M., & Shinn, M. R. (2002). Structuring school-based interventions to achieve integrated primary, secondary, and tertiary prevention goals for safe and effective schools. *Interventions for academic and behavior problems 2: Preventive and remedial approaches,* 1–25.

Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H. H., & Feil, E. G. (1998). First step to success an early intervention approach for preventing school antisocial behavior. *Journal of Emotional and Behavioral Disorders, 6*(2), 66–80.

Walker, H. M., Ramsey, E., & Gresham, R. M. (2004). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont: Wadsworth.

Walker, B., Cheney, D., Stage, S., Blum, C., & Horner, R. H. (2005). Schoolwide screening and positive behavior supports: Identifying and supporting students at risk for school failure. *Journal of Positive Behavior Interventions, 7*(4), 194–204.

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavior Disorders, 13*(2), 79–96.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*(2), 141.

Wills, H., Greenwood, C. R., Kamps, D., Heitzman-Powell, L., & Selig, J. (2010). The combined effects of grade retention and targeted small-group intervention on students' literacy outcomes. *Reading & Writing Quarterly, 26*(1), 4–25.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding it's heart. *Journal of applied behavior analysis, 11*(2), 203–214.

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology* (Vol. 4, pp. 251–288). Mahwah: Erlbaum.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49,* 156–167.

Implementing Response to Intervention in Secondary Schools

Mark R. Shinn, Holly S. Windram and Kerry A. Bollman

This chapter addresses key elements in secondary-level (i.e., middle and high school) implementation of response to intervention (RTI), more suitably labeled multi-tiered system of supports (MTSS). The label MTSS is increasingly preferred because many educators interpret RTI narrowly as just the process for determining eligibility for special education under the category of specific learning disabilities (SLD). For many secondary educators, in settings where SLD identification occurs much less frequently than in elementary schools, this narrow RTI interpretation understandably diminishes their interest and engagement. SLD eligibility has little to do with their day-to-day work. An MTSS perspective, however, enables a *whole school* and *all teachers* school improvement perspective by increasing attention to improving the quality of academic and behavior support to meet the needs of *all* students, preventively and by providing increasingly intensive basic skills remediation to those students who still need it. Furthermore, an MTSS perspective fosters necessary attention towards evidence-based academic and behavioral support strategies necessary to achieve the desired outcomes of the *Common Core State Standards* (*CCSS*; National Governors Association Center for Best Practices and Council of Chief State School Officers 2012)

This chapter provides a specification of what features and actions of secondary MTSS are similar to elementary-level practices and, most importantly, what features and actions are *different* and require special attention for implementation success. For example, one similarity is a *prevention* and *early intervention focus* driven by systematic screening. However, secondary screening implementation differs with an increasing shift from universal screening (i.e., testing all students) in early middle school to individual, case-by-case screening by high school. Another similarity is a strong emphasis on research-based tier 1 instructional practices. However, secondary MTSS shifts from the elementary emphasis on basic skills to an increased focus on content-area curricula such as science, social studies, advanced mathematics, and students' skills in navigating complex informational text and writing using evidence. Finally, among these consistencies with elementary MTSS, secondary implementation also values a more positive school climate and educative positive behavior support practices with increasingly intensive interventions. However, secondary MTSS behavior support practices, shift attention to prevention and intervention for tardies, truancy, and more effective

M. R. Shinn (✉)
National Louis University, Chicago, USA
e-mail: markshinn@icloud.com

H. S. Windram
Hope Network, Michigan Education Corps,
Grand Rapids, USA

K. A. Bollman
TIES, Cleveland Heights, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_32

discipline consequences than typical practices of detention, suspension, or expulsion (Sprick and Borgmeier 2010).

It is widely recognized that secondary MTSS implementation lags behind. In the last nationally reported survey (Spectrum K12 School Solutions 2011), 51 % of over 1000 school district administrators indicated that they had implemented MTSS in their high schools. This figure represented a 16 % increase from 2010. However, these figures are open to skepticism expressed in a number of resources. For example, as part of the National RTI Action Network, Burns et al. (2013) noted that only 18 % of their secondary respondents indicated that they used tiered reading services, typically by far the most common intervention area. Another recent source (Prewett et al. 2012) also suggested that Spectrum K–12 survey implementation rates may be inaccurate. They characterized middle school implementation as "relatively rare" (p. 137), a perspective echoed by Duffy's judgment (2007) that few high schools have systematically implemented tiered interventions.

Why is Secondary Implementation Lower and Slower?

This chapter posits that the biggest single barrier to secondary MTSS implementation is a *confused or unclear purpose* that inadvertently fails to engage secondary educators at best and under some circumstances can actively encourage strong disinterest. Without a clear MTSS purpose, it is easy to extend elementary MTSS practices to secondary schools without considering the developmental educational needs of older students and the culture, climate, and training of secondary teachers, nearly all of whom are content-area specialists. If the primary purpose of secondary MTSS is interpreted solely to increase content-area achievement success, then schools may attempt to develop tiered assessment and instruction of increasing intensity around specific course content like social studies or science. It follows that special education eligibility for SLD may be determined on the basis of having a social studies achievement discrepancy and a failure to benefit from tiered social studies intervention(s).

This focus is illogical and impractical to secondary educators. Far too many students who do not have disabilities struggle with content course instruction (Schumaker and Deshler 2010). Further, offering tiered versions of content-area courses like science and social studies presents unreasonable logistical and structural challenges like staffing, scheduling, and meeting graduation requirements. In addition, the potential for tracking students into lower-level courses with diminished expectations and lowered instructional quality is increased.

Regarding assessment, few, if any, validated content-area screening and frequent progress monitoring tests exist. All the screening tests reviewed for use in RTI by the US Department of Education National RTI Center (www.rti4success.org) are in the areas of basic skills. Furthermore, despite some success in algebra (Foegen 2012) and promising research efforts by Espin (e.g., Espin and Campbell 2012; Espin and Tindal 1998) and McMaster (e.g., McMaster and Espin 2007; Wallace et al. 207), no other frequent (i.e., at least monthly) content-area progress monitoring tools have been validated.

Another common misinterpretation of MTSS in secondary settings is seeing the primary purpose as extending Tier 1 elementary implementation practices to middle and high schools (i.e., focus on significantly increasing reading, writing, and mathematics basic skill acquisition). In this scenario, schools may put the burden of providing intensive basic skills interventions onto the already full plate of content-area teachers who lack the time and training for this remediation. Prewett et al. (2012) documented this extension approach, stating "*recently* (our emphasis) middle schools began adopting RTI frameworks largely based on elementary to secondary practices generalization" (p. 136).

This generalization seems rational given the origins of early intervening services through a problem-solving model in the early 1980s (e.g., Marston and Magnusson 1985; Tindal et al.

1985), and rapid acceleration of MTSS with a K–3 reading focus following No Child Left Behind (2001).[1] In reality, intensive teaching of basic skills makes little sense to secondary teachers who already struggle with many students who lack a range of basic prerequisite study and academic skills to benefit from more complex content-area instruction. Training and time issues aside, it is a difficult "sell" to get a grade 8 science teacher to accept responsibility for teaching intensive basic reading, writing, and mathematics skills *and* science to students who lag behind.

Furthermore, it also makes little sense to secondary teachers and administrators to continue elementary assessment practices like *benchmarking* (i.e., basic skills testing students three times per year for universal screening *and* universal progress monitoring). With each increasing grade beyond grade 6 in most communities, these *benchmark* data are not useful for progress monitoring, when nearly all students already have sufficient basic skills and the repeated testing is inefficient and uneconomical in terms of instructional time and money. The most significantly discrepant students already have been identified at younger grades. By grade 7, nearly all students reach a ceiling and demonstrate an asymptotic trend of growth in their basic skills improvement (Fuchs et al. 2001; Jenkins and Jewell 1993).

Typical students show little growth on the commonly used basic skills tests. Consequently, basic skill progress monitoring included as part of *benchmark* assessment also is inefficient and uneconomical and, without clear return, turns off secondary teachers as simply a matter of procedural compliance associated negatively with MTSS. *Overall, we agree with our secondary educator contemporaries that the aforementioned applications of MTSS at the secondary level would be misguided and insensible allocation of resources like staffing and instructional time.*

[1] See Germann (2010) for a history of the former and Gresham, Reschly, and Shinn, (2010) for a chronology of the latter.

Purpose of the Chapter

This chapter proposes that the appropriate purpose of secondary MTSS is to advance college and career readiness in two important ways: (a) to increase the quality and quantity of evidence-based instructional and behavioral support practices in core content-area instruction *and* (b) to enable basic academic and behavioral skills interventions of suitable intensity to be provided to those students who remain discrepant. In attempting to communicate some of the essential features of secondary MTSS implementation, it should be noted that this chapter cannot comprehensively describe each and every component that may be valuable to student achievement and behavioral success. However, efforts will be made to identify additional readings and references to support greater understanding.

What Is in it for Secondary Teachers: Real Life Scenarios with MTSS Solutions

The optimistic perspective in secondary MTSS implementation is that validated, effective, and powerful practices have been identified that are robust across secondary content (e.g., Coyne et al. 2007b; Hattie 2009; Marzano 2007). Within specific secondary content areas, instructional practices and content with strong research support also have been identified including adolescent literacy (Kamil et al. 2008; Torgesen et al. 2007a, b), writing (Coyne et al. 2007a; Harris et al. 2008; Troia and Olinghouse 2013), mathematics (Gersten et al. 2009; National Mathematics Advisory Panel 2008; Thompson 2006), science (Grossen et al. 1994), and social studies (Carnine et al. 1994; Carnine et al. 2007). However, persons with knowledge of research-based instructional and behavioral support strategies who visit any middle or high school will find many instances where there is a clear research-to-practice gap. That is, many practices *with* strong empirical support (e.g., frequent formative assessment, use of graphic organizers) are not used very often. Conversely, many practices with *little* empirical support (e.g., summarization, highlighting, and underlining) are routine.

Acknowledging this gap is not intended to be critical or blame secondary educators who passionately work to meet the needs of diverse adolescents who come to classrooms with a range of skills and interests. As noted by Dunlosky et al. (2013) "some effective techniques are underutilized—many teachers do not learn about them, and hence many students do not use them, despite evidence suggesting that the techniques could benefit student achievement with little added effort" (p. 5).

The gap between research and practice is not limited to academic practices. A similar and longstanding gap exists between school discipline practices in and around the school building as well as in the classroom (Frey et al. 2010; Sprick and Borgmeier 2010). Instead of a proactive, preventive approach, middle and high school discipline practices can still be characterized as relying heavily on the traditional, albeit ineffective, practices of detention, suspension, and expulsion within a tradition of zero tolerance (Cook et al. 2008; Sprague and Walker 2010).

The following four scenarios illustrate common research-to-practice gap examples and how MTSS can contribute solutions to meet these challenges:(a) a science teacher who values vocabulary, (b) grade 9 universal reading screening results in a diverse high school showing the range of students' reading skills, including students with severe achievement needs, (c) an observational study of high school special education instruction, and (d) the office referral and tardiness rates in an urban middle school.

MTSS Can Increase Attention to Research-Based Practices in Content-Area Classes

In this scenario, a ninth-grade science teacher is aware that specific science vocabulary knowledge is related to content-area literacy and achievement success. These science vocabulary words are unlikely to be learned incidentally, even in other science courses or by wide reading. These discipline-specific words would be considered tier 3 words, not to be confused with tier 3 intervention, by Beck et al. (2002). The science textbook identifies critical science-specific words such as *alkali metals* and *metalloids* in a unit on elements. The science teacher *may* be aware that among the ineffective ways for students to learn these vocabulary words would be to have students rely on the science text's glossary (Baumann and Kame'enui 2004; Kame'enui and Baumann 2004). Equally ineffective would be expecting students to look up the definition of unknown science words in the dictionary. Instead, to increase science achievement, these words need to be *taught* to students. Therefore, the science teacher spends planning time creating weekly handouts for students with the critical vocabulary words identified with written definitions on Monday. Students also are provided access to online versions of the vocabulary word lists. Examples from three of 21 weekly words are presented in Table 1.

Because the teacher values vocabulary, the students are tested at the end of the week with a quiz where students are expected to *write* the definitions rather than just match the definition to the word.

Across the 21 science words, the science teacher provided 441 total words in the definitions and students are expected to memorize them, a process that is described as learning *independent member sets* (Engelmann and Carnine 1982). This process requires students to be motivated *and* memorize the definitions one by one. The implicit emphasis is on each specific science word, not the components of the definition.

Historically, some of the class's science students may lack motivation and/or the memorization strategies to learn these vocabulary words and their achievement would suffer. Some of these students may have even failed science and been referred for, and placed in, special education as SLD. MTSS implementation could provide this well-intentioned science teacher with staff development and support to use more effective, evidence-based instructional techniques, including vocabulary instruction strategies.

Table 1 Sample grade 8 science vocabulary words from the 21 weekly words

1. Alkali metals (e.g., lithium, sodium, potassium)	Physical: soft, silvery, white in appearance, good conductors of heat/electricity
	Chemical: most reactive metal, always combined in nature, reacts violently with water, have 1 valence electron, loses one electron when bonding, usually bond with the halogens
2. Alkaline earth metals (e.g., magnesium, calcium, barium, strontium)	Physical: similar to alkali metals, good conductors of heat/electricity
	Chemical: reactive, not as reactive as alkalis, always combined in nature, have two valence electrons, loses two electrons when bonding
3. Noble gases (e.g., helium, neon, argon, krypton, radon)	Physical: all gases at room temperature
	Chemical: usually unreactive, rarely combined in nature, valence shell is full, no need to bond since the atoms of these elements have their octet (full outer shell) already

Considerable evidence has been accumulated about what works to teach vocabulary (Kame'enui and Baumann 2004; Stahl and Kapinus 2001) and increase content-area reading success (Torgesen et al. 2007a, b). These research-based explicit vocabulary instruction strategies have been shown to increase achievement not only for struggling students but on average, to increase achievement for all students (Torgesen et al. 2007a, b).

In secondary MTSS implementation, the plan would be to provide an evidence-based vocabulary that could be used by *all* content-area teachers such as the *LINCS vocabulary strategy* (Schumaker and Deshler 2010). When students are expected to learn a common learning strategy that is employed consistently across content classes, they achieve mastery in its use. Alternately, MTSS could support the use of a graphic organizer to make the implicit vocabulary word attributes explicit that are the same or different across the 21 vocabulary words. See Table 2 for an example of how the three sample words could be organized by making the implicit attributes of the definitions more explicit.

By making the implicit definitional attributes (e.g., appearance, conductivity) more explicit, it is possible to reduce the disadvantages of students learning independent member sets (e.g., memorization) and enable other organizational strategies. For example, a student could organize the elements by appearance (e.g., which elements are soft and silvery?) or by bonding characteristics (e.g., which elements have no need to bond?) or compare and contrast strategies to organize information. MTSS can enable increased implementation of evidence-based practices to increase science achievement for more students, reduce failure rates, and decrease the need for special education.

MTSS Can Increase the Range of Powerful Interventions for Students Who Need Them

Unfortunately, for some secondary students, even use of high-quality content instructional practices such as explicit vocabulary instruction cannot meet their content learning needs simply because they lack the sufficient level of basic skills to even be able to read the science text. Content-area understanding is elusive under these circumstances. Far too many students experience reading difficulties that can impede their content-area classes' success. The most recent National Assessment of Educational Performance (NAEP) report found that although reading improvement over time has been noted, nearly one in four (22 %) eighth graders scored below *basic*; almost two thirds (64 %) were judged as below *proficient* (National Center for Education Statistics 2013). A similar pattern of below standards-level performance was reported in mathematics with 26 % of eighth graders scoring below *basic* and 64 % below *proficient*.

Table 2 How sample grade 8 science vocabulary words could be organized and displayed

		Physical features		Chemical features			
Element	Examples	Appearance	Conductivity	Reactivity	Electrons	Combines	Bonds
Alkali	Lithium, sodium, potassium	Soft, silvery, white	Good with heat, electricity	*Most* reactive	−1 valence	*Always* with nature	Usually with halogens
Alkaline	Magnesium, calcium, barium, strontium	Soft, silvery, white	Good with heat, electricity	Reactive, but *not* as reactive as alkalis	−2 valence	*Always* with nature	
Noble gas	Helium, argon, neon, krypton, radon	Invisible. All gases at room temperature		Usually unreactive	Octet (full outer shell already)	*Rarely* with nature	No need to bond

A real-life example of grade 9 students' range of reading scores is shown in Fig. 1. In this diverse high school, all entering grade 9 students took the *Scholastic Reading Inventory* (SRI; Scholastic Inc. 2001). Students were tested and predicted Lexile scores were obtained that identified the level of text they would have success reading and understanding. Results showed that sizable numbers of students were predicted to have considerable difficulty understanding grade 9 text. For example, 25 students obtained predicted Lexile scores of grade 6 proficiency, 3 years below expectations. Another 123 students performed even lower. With almost 150 students (26 %) significantly discrepant from the expected grade-level reading skills for content-area success, even evidence-based content-area instructional practices such as those illustrated in the previous scenario are not likely to produce the desired student achievement outcomes. These students need to increase their basic reading skills significantly. Special education resources alone, even if sufficiently staffed to provide the intensive reading instruction these students require, would not meet the demand.

MTSS implementation could provide this school a range of appropriately intensive interventions aligned with students' reading

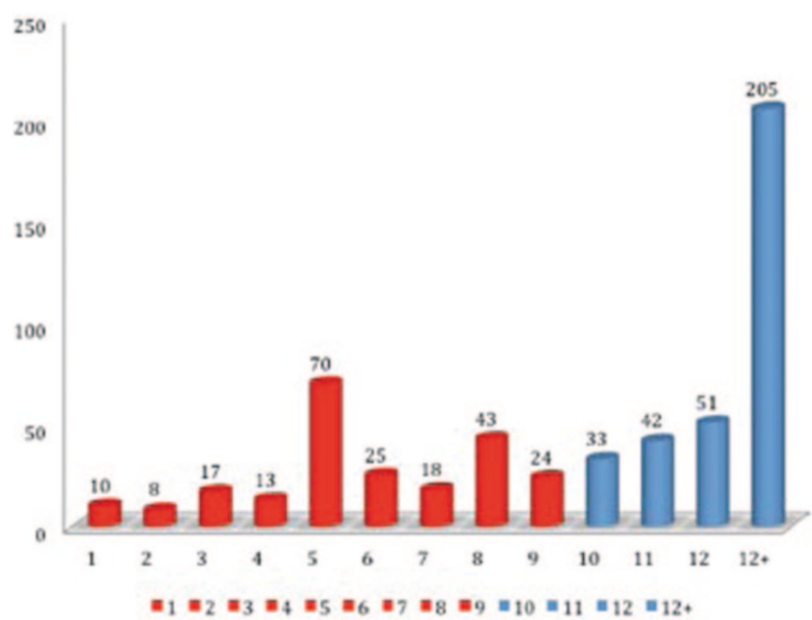


Fig. 1 A frequency distribution of grade 9 Lexile proficiencies predicted by the SRI. SRI scholastic reading inventory

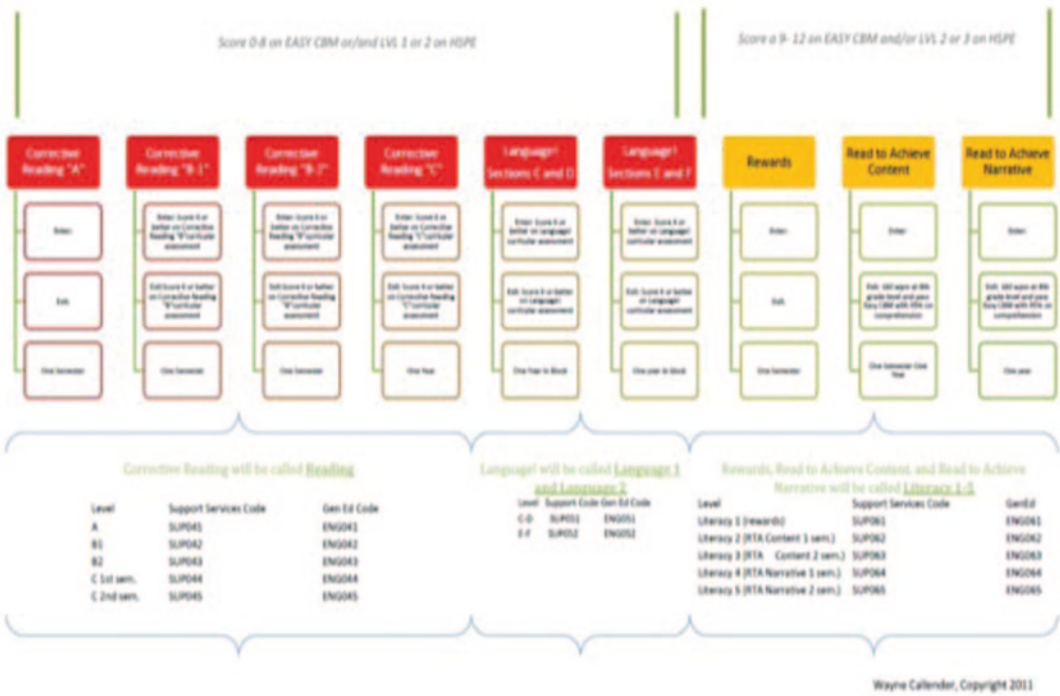


Fig. 2 A diverse high school's range of increasingly intensive intervention offerings triaged by reading screening results. From Wayne Callender, Ph.D. Partners for learning, http://partnersforlearning.org

needs. One high school's response to reading skill diversity enabled through MTSS implementation is shown in Fig. 2. All high school students were screened using *easyCBM* (Houghton Mifflin Harcourt Riverside Assessment, 2013), a set of short tests based on curriculum-based measurement (CBM) test development strategies (Jenkins and Fuchs, 2012).

The high school's MTSS plan enabled evidence-based reading interventions of increasing intensity to be staffed and scheduled as regular courses based on students' reading screening results. For example, students scoring 9–12 on *easyCBM* were triaged into either a reading intervention using *REWARDS* (Archer and Gleason, 2001), a program designed to address multi-syllabic words, vocabulary, and comprehension strategy deficits. Students with higher reading skills were triaged into classes where either *Read to Achieve Narrative* or *Read to Achieve Content* (Marchand-Martella and Martella 2010) that emphasized evidence-based strategies to navigate specific text structures were used. Increasing intensive interventions reaching as low as nonreading high school students using SRA Corrective Reading (Engelmann et al. 1999) were scheduled, providing a full range of reading intervention options.

MTSS Can Increase Special Education Quality

When secondary students are significantly discrepant in their achievement, the historical solution has been special education. With so much of the emphasis on the quality of elementary special education, the question of what is and how effective secondary special education is has been largely ignored in the professional literature. Almost a quarter of a century ago, Zigmond (1990) questioned whether the special education services students received were aligned with their instructional needs. Secondary special education students with similar achievement needs could be served through a variety of different approaches, including, but not limited to full-inclusion, co-

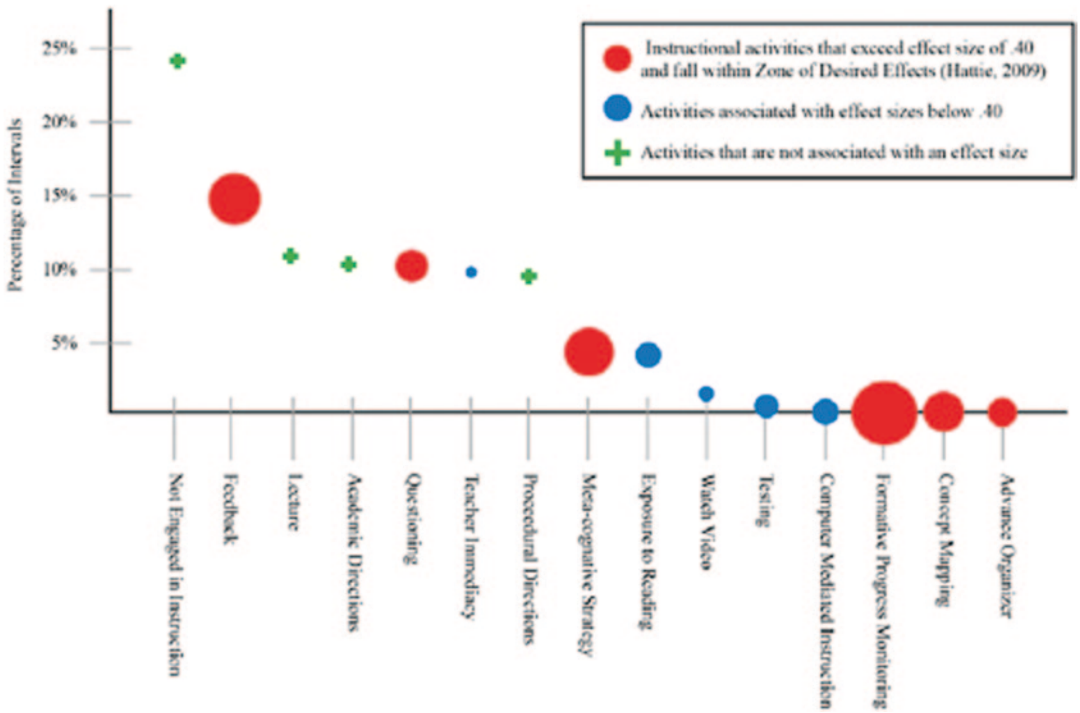


Fig. 3 Frequency of observed special education instructional practices by effect sizes. From Deshler and Cornett (2012)

teaching, push-in services, or pull-out resource programs. The instructional focus also varied from supporting content-area (e.g., social studies, science) achievement to help with homework, alternative assessments and expectations, and modified grades. Almost 20 years later, Conderman and Petersen (2007) expressed concern that with increasing standards, little has changed in terms of providing instructionally aligned interventions to secondary special education students, echoing concerns over the "tutoring trap" (Deshler et al. 1996; Deshler and Schumaker 2005) or special education teachers reporting feeling they are used like instructional aids. Unfortunately, too often, the special education services secondary students receive are not aligned with their instructional needs. Students who need intensive reading instruction are as likely to not receive any special education reading program as they are to receive reading instruction at all (Phillips 2002).

The quality and quantity of instructional practices secondary special education students receive also is of concern (Deshler and Cornett 2012). These authors observed instruction in three settings, (a) special regular classrooms, where a special education teacher delivered content-area instruction (e.g., social studies); (b) co-taught content classes with a general education teacher assuming the primary instructional responsibility, and (c) special education resource classrooms. Instructional practices that were employed were recorded and categorized by the magnitude of the impact on student achievement by effect size as reported in Hattie's extensive meta-analysis (Hattie 2009). Results are shown in Fig. 3.

Few instructional practices associated with meaningful achievement gains (i.e., effect sizes greater than 0.40) were observed and most of the instructional practices recorded were *not* associated with student gains and likely contributed to the high rates of high school dropout.

MTSS redresses the problem of quality secondary special education services by providing schools a clearer focus on who receives special education (Germann 2010, 2012) and by expanding the range of evidence-based practices that can increase student achievement (Deshler and

Cornett 2012; Lenz et al. 2003; Schumaker and Deshler, 2010).

MTSS Can Increase the Quality of Behavior Support

The final scenario is from a diverse middle school of more than 1500 students where more than 90% of students qualified for free or reduced lunch. Like many secondary schools, school administrators and staff were being bogged down with an average of 250 tardies a day, many of which resulted in office referrals, totaling almost 3200 for the year. With an estimated staff time cost of 15 min for each office referral (Sprick and Borgmeier 2010; Sprick and Garrison 2008), approximately 800 hours of educator time was expended, likely with little impact on student behavior. MTSS has the potential to redress the ineffective and inadequate responses to challenging and inappropriate behavior by enabling attention to systematic school- and class-wide evidence-based positive behavior supports. In this school's case, school staff were trained to use START On Time (Sprick 2006), where teachers worked in teams to acknowledge positive attainment of class start times and conduct regular "sweeps" for tardy students. Results showed a 70% daily decrease in tardies and a 67% decrease in office referrals. These results, when positive behavior approaches are implemented through MTSS, are not unusual. See, for example, the outcomes reported by Schuta et al. (2012) and Ward and Gersten (2013).

Critical Underpinnings of Secondary MTSS Implementation Success

When retracing the history of MTSS/RTI, Germann (2010, 2012) discussed the challenges presented to secondary special educators in providing better services and meeting the needs of more students through general education improvements. He described existing efforts as maintaining a system of socially sanctioned segregation by removing the student from content-area classes, if not expectations, and providing the wrong solution to the wrong problem. That is, at middle and high school, students are often referred for special education because they were failing in content-area classes (e.g., social studies, English). Once they are determined eligible, students receive interventions that are not particularly powerful, whether they are accommodations like extended time or modified grades, or interventions that provide, at best, only short-term solutions, such as help with homework or alternative content courses (Conderman and Petersen 2007; Prewett et al. 2012; Schumaker and Deshler 2010). In only rare instances are the necessary interventions provided that can provide short- and long-term solutions, like intensive basic skill remediation and robust content-learning strategies. Secondary MTSS cannot flourish under these conditions.

To increase the likelihood of secondary MTSS success, four major tasks must be accomplished that relate directly and indirectly to special education:

1. The problems to be addressed by MTSS remedial services must be linked to a *dual-discrepancy model*, clearly defining the problem for intervention.
2. SLD eligibility criteria must be changed to align with the dual-discrepancy model when students fail to benefit from appropriately intensive intervention, differentiating those students who receive *treatment* from those who require *support*.
3. Special education interventions must be focused primarily on ensuring robust learning strategies and/or intensive, research-based basic skills interventions; and
4. Special education IEP goals and progress monitoring practices must reflect research-based IEP and progress monitoring practices.

Embracing the Dual-Discrepancy Model As the Core of Secondary MTSS Design

At the elementary level, the *dual discrepancy model* (Fuchs and Vaughn 2005; Pericola et al. 2003) implicitly underscores tiered-services implementation. As noted throughout this book (see Chaps. X, Y), *screening* identifies students with *performance discrepancies*, those students significantly below a normative or standards-based level of basic skills performance. Other chapters in this book also have presented descriptions of how elementary level progress monitoring is conducted (see Chaps. X, Y) to measure *progress discrepancies,* when students fail to attain an expected rate of improvement (ROI).

Curriculum-based measurement (CBM), a set of short, scientifically sound tests of basic skills (Deno 1985, 2012; Shinn 1989; Shinn 1998) is commonly used in elementary MTSS to measure these discrepancies. CBM is a generic term for a broad "family" of measures reflecting general outcome measurement (GOM; Fuchs and Deno 1991; Jenkins and Fuchs 2012) test design principles. These short tests represent *broad basic skills constructs* (e.g., general reading achievement, mathematics computation skills) rather than specific skills (e.g., identifying an author's purpose, solving quadratic equations).

A number of different publishers provide CBM tests, including, but not limited to *aimsweb, easyCBM, DIBELS*, *Yearly Progress Pro, and FAST*. All but *FAST* have been reviewed independently by the National Center for Response to Intervention (www.rti4success.org) and the National Center for Intensive Intervention (www.intensiveintervention.org) for screening *and* progress monitoring use. *FAST* has been reviewed for progress monitoring by the National Center for Intensive Intervention.

Some schools use the computer-delivered tests *NWEA Measures of Academic Progress (NWEA* 2014), or *STAR* assessments (Renaissance Learning 2012) in MTSS. The former has been reviewed for screening by the National Center for Response to Intervention and the latter by both centers.

Depending on test publisher, screening and progress monitoring tests are available for language arts (i.e., reading, written expression, spelling) and mathematics. The most frequently used measure for determining the reading performance and progress discrepancies is Reading-CBM (R-CBM). This test is a short measure of general reading skill where students read a passage aloud for 1 min and the number of words read correctly are counted. It is especially useful for frequent progress monitoring due to its sensitivity to improvement in short periods of time (e.g., 4–6 weeks).

R-CBM use is not limited to elementary RTI and is equally appropriate at secondary. For example, Denton et al. (2011) and Espin et al. (2010) found correlations among R-CBM and other measures of reading sufficient to provide validity to conclusions about the general reading skills of middle school readers. Barth et al. (2012) demonstrated that both median and mean scores are reliable and have convergent validity for both typically developing and struggling middle school readers.

Maze, a group or individually administered silent reading test, also is used to assess general reading skills and also has demonstrated criterion-related validity with secondary populations (Torgesen et. al. 2005; Espin and Foegen 1996; Tolar et. al. 2012). It offers schools the capacity to efficiently screen for reading performance discrepancies.

The Dual Discrepancy Model in Middle School At middle school, the same (a) basic skills achievement screening measures (e.g., CBM), and (b) criteria/cut scores for identification/early intervention are used the same way as at the elementary level to determine the performance discrepancy. For example, when R-CBM is used, students who *perform below average* (i.e., the 25th percentile, a *norm-based* approach) or *are predicted to not pass the next high stakes state test* (i.e., a *standards-based* approach) are identified for appropriately intensive intervention (Shinn 2010).

A screening example of a middle school grade 8 student with a significant norm-based reading performance discrepancy using CBM is shown in Fig. 4.

When given three randomly selected grade 8 reading passages, this student read 80 words correctly (WRC) during the school's *universal* screening (i.e., *all* grade 8 students are tested). This score would place the student well below

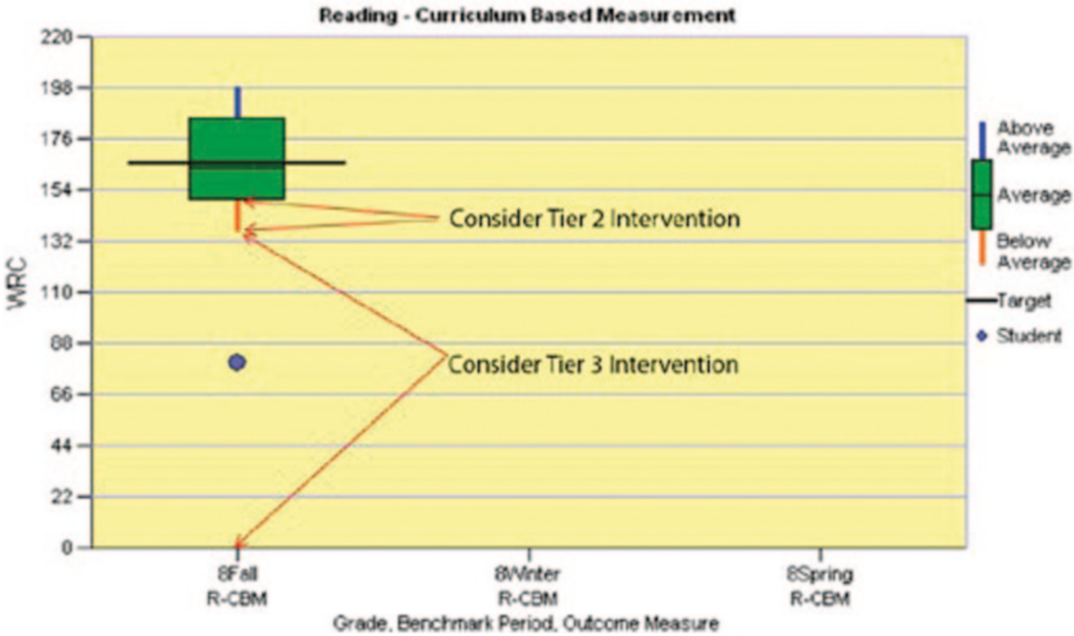


Fig. 4 A grade 8 student's screening results showing a significant reading performance discrepancy that may require tier 3 basic skills reading intervention

the 10th percentile of other grade 8 students, a common criterion suggesting the need for tier 3 intervention.

Although the middle school screening achievement *measures* typically are the same as elementary, the assessment *practices* to determine the performance discrepancy differ. Instead of the *repeated* universal screening practices within a grade accomplished with benchmark assessment process (i.e., three times per year), at an agreed upon grade level (e.g., grade 7) screening shifts from benchmarking to *once a year* universal screening. It is suggested that this screening occur at the *end* of the academic year so that students can be identified for potential intervention for the following academic year. This timeframe enables appropriately intensive interventions to be built into students' original schedules, avoiding the considerable logistic challenges that within-year schedule changes present. Additionally, students who may need intensive intervention can be identified for potential summer school remediation.

When determining potential progress discrepancies, the middle school progress monitoring *measures* again are the same as elementary, but the assessment *practices* to determine the progress discrepancy differ at tier 1. At the elementary level, tier 1 progress monitoring occurs with a benchmark assessment approach. All students' rate of improvement from fall to winter to spring is determined. At middle school, when benchmark assessment shifts to once per year universal screening, tier 1 basic skills progress monitoring for all students is discontinued. However, the frequent basic skills progress monitoring practices at tiers 2, ranging from once per month to weekly, and weekly tier 3 testing remain the same. An example of how the progress discrepancy is measured for a student receiving a tier 3 intensive basic skills reading intervention is shown in Fig. 5.

A single R-CBM sample from grade-level reading passages is collected weekly and progress is evaluated against an expected rate of progress (i.e., the aim line or solid line in the figure) that reduces the performance discrepancy. Like all formative assessment practices, intervention is adjusted when progress is unsatisfactory.

Fig. 5 Measuring a grade 8 student's reading progress discrepancy

The Dual Discrepancy Model in High School In high school, screening to determine the performance discrepancy differs from elementary practices in three ways. First, the shift from benchmark assessment to once per year universal screening occurs no later than at grade 9 in typically achieving schools. Second, at a predetermined point (e.g., grade 10), universal screening shifts from universal to *individual screening*. With respect to identifying severe basic skill deficits, by high school, nearly all students should have been identified and the expenditure of time and resources to test all students to find any unidentified ones makes little sense. Only individual students with potential basic skills performance discrepancies are screened.

Third, and most importantly, what is most notably different from both elementary *and* middle school practices is the explicit shift in *how the performance discrepancy is defined*. Fundamental to this process is a *standards-based criterion*, establishing a minimum performance standard below which the community has agreed that intensive intervention is required (Shinn 2008; Tindal et al. 1987; Tindal et al. 1985). One common minimum standard is an end-of-grade 7 reading proficiency. This standard can be operationalized by using the normative performance of end-of-grade 7 students. Any individual high school student who reads below this standard could be considered for intensive basic skills intervention.

An example of a grade 11 student with a significant standards-based reading performance discrepancy using CBM is shown in Fig. 6.

This grade 11 student had moved into the high school from a different community and within a few weeks, was identified as struggling with multiple content-area classes. Individual screening was conducted to rule out a potential severe basic skills performance discrepancy. When given three randomly selected grade 7 reading passages, representing the *minimum* reading standard, the student read 80 words correctly per minute (WRC). If the student's scores were in the "box" or higher in the box-and-whisker chart, the student would not be judged to have a severe performance discrepancy. In this example, the individual screening score (the dot on the figure) placed the student well below the average range of grade 7 students in the community, suggesting the need for intensive tier 3 intervention.

The frequent progress monitoring practices at tiers 2 and 3 to determine the *progress discrepancy* remains the same as elementary and middle school in terms of how often progress is measured. However, the criterion for acceptable performance (CAP) also shifts from a norms-based to a standards-based criterion. That is, when the performance discrepancy is defined as below an end-of-grade 7 reading standard, the progress monitoring goal would be to have the student

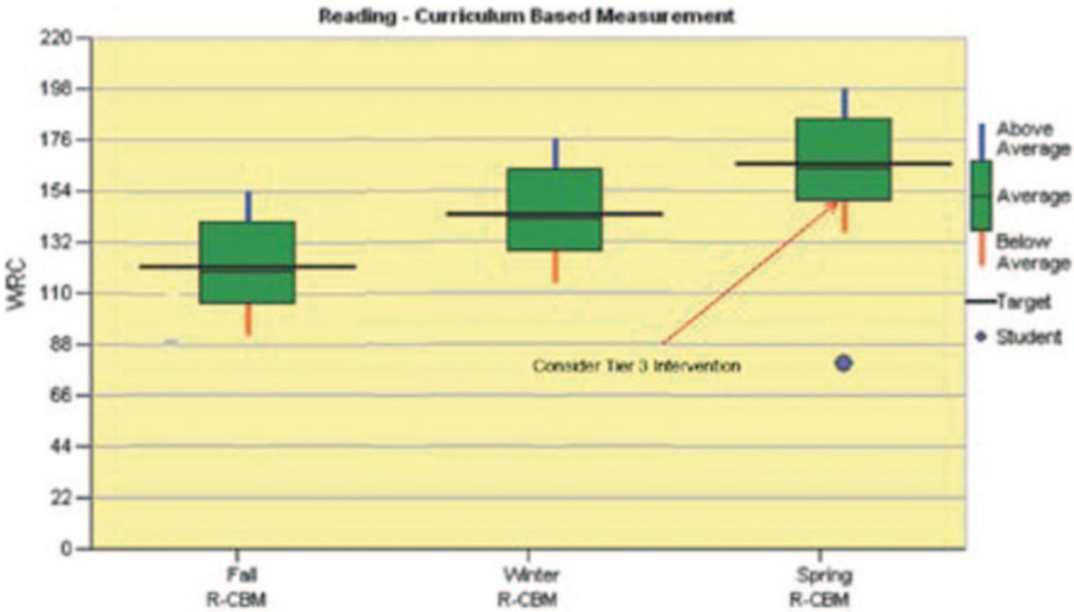


Fig. 6 A grade 11 student's individual screening results showing a significant reading performance discrepancy that may require tier 3 basic skills reading intervention

read as well as an average end-of-grade 7 student by the end of the intervention period. Student progress would be measured weekly using grade 7 reading tests.

A summary of screening and progress monitoring practices to determine performance and progress discrepancies by middle and high school grades and tiers is shown in Table 3.

Establishing the Dual Discrepancy As the Basis for SLD Eligibility

Students who struggle with content-area achievement in middle and high school and/or behavior form the basis for nearly all new special education referrals (Schumaker and Deshler 2010; Shinn 2008). However, if a content-area performance discrepancy (e.g., discrepant social studies achievement and/or failing courses) drives special education eligibility, the "wrong" interventions will be delivered to the "wrong" students with little discernable benefits (Walker 2004; Walker and Shinn 2010). That is, as discussed earlier in this chapter, these students too often receive content-area tutoring, help with homework, and/or accommodations rather than the intensive basic skills interventions and evidence-based learning strategies (Deshler and Schumaker 2005; Schumaker and Deshler 2010)

Table 3 Screening and progress monitoring practices summary

Grades	Performance discrepancy	Tier 1 progress	Tier 2	Tier 3
[a]6	Benchmarking	Benchmarking	Monthly, every other week, or weekly	Weekly
7–9	End-of-year universal screening	Not applicable	Monthly, every other week, or weekly	Weekly
10–12	Individual Screening	Not applicable	Monthly, every other week, or weekly	Weekly

[a] Use grade 6 spring benchmark assessment for grade 7 tier 2, tier 3 intervention planning

Table 4 A set of strategies and criterion for determining specific learning disabilities special education eligibility in middle and high school

Students may be eligible for special education under the category of SLD if:
1. Severe performance (achievement) discrepancy where students perform below the 10th percentile of local peers determined by CBM grade-level screening at grades 6–8, or Severe achievement (achievement) discrepancy where grade 9–12 students perform below the median of local end-of-year grade 7 students on CBM grade 7 screening tests
2. Severe progress discrepancy where the actual rate of improvement (ROI) is below the ROI that significantly reduces the severe performance discrepancy
3. When tier 3 intervention is a research-based, basic skills intervention of appropriate intensity;
4. Observation documents that the tier 3 intervention is delivered with fidelity
5. The proposed special education intervention has a direct instruction, basic skills focus that is described in sufficient detail to suggest that it is different in meaningful ways from tier 3 intervention and reflects specially designed instruction to meet the student's unique needs; and
6. All other IDEA procedural requirements (determinant and exclusionary factors) have been addressed

they may require to benefit from middle and high school instruction.

The cornerstone of RTI as a component of SLD eligibility determination has been the *dual discrepancy* discussed throughout this book and illustrated in the previous section of this chapter. To ensure MTSS implementation success, adoption of the middle and high school strategies to operationalize special education eligibility using the performance and progress discrepancies is essential. One way that SLD eligibility can be determined at middle and high school is operationalized in Table 4.

Once changes in special education eligibility using an RTI and dual-discrepancy approach are in place, the opportunity for a fundamental shift in school culture and changes in intervention services is created. That is, schools can shift their culture by clarifying *who* they serve in special education and *what* services they provide to them. No longer is every student struggling with content curriculum potentially eligible for special education where they are provided interventions that lack impact. Instead, middle and high schools can shift to identifying those students with *severe basic skills performance discrepancies* they can *treat* early and with powerful remedial interventions. Special education students who require treatment are delivered an intensive basic skills intervention using validated learning strategies in the areas of need, enabling special education to deliver powerful evidence-based programs that may reduce performance gaps.

Schools then are in a better position to differentiate these students from others students with minimal level basic skills that require *support* in content-area courses. This support requires all teachers to learn and use evidence-based practices (e.g., teaching routines) with the further benefit that these strategies have been shown to increase student achievement for a broad range of students' skills (e.g., Kamil et al. 2008; Lenz et al. 2003; Lenz et al. 2005; Torgesen et al. 2007a, b).

Providing the Intensive Basic Skill Interventions and Learning Strategies Students Need

A consistent theme of quality middle and high school MTSS/RTI has been an alignment of intervention services with the types and severity of problems that students display. Yet, as often as not, special education services have been misaligned. As stated by Conderman and Petersen (2005) "many secondary special educators may experience increased pressure from parents, administrators, general educators, and students to provide homework assistance and review or reteach content-area subject matter" (p. 235). In all likelihood, increased use of one particular special education service delivery approach, co-teaching, may inadvertently sustain this tutoring trap. Without specialized training, special education teachers co-teaching students jointly with content-area teachers in content-area classes seem very likely to result in the same prevailing and ineffective practices.

Another factor compounding the misalignment of instruction with student need is that schools commonly overlook the IDEA-2004 requirement that IEPs contain a statement of special education services be based on peer-reviewed research to the extent practical (Rose and Zirkel 2007). As noted earlier in the chapter and described by Deshler and Cornett (2012), secondary special education instruction can be characterized by a lack of practices supported by peer-reviewed research.

MTSS offers two compelling, research-based intervention alternatives for special education students with reading performance discrepancies. First, co-teaching with a particular set of strategies, the Strategic Instruction Model Content Literacy Continuum (CLC), and second, explicit, research-based program-driven intensive basic skills instruction delivered in scheduled intervention classes.

Content Literacy Continuum (CLC) Co-teaching with CLC is an excellent way to deliver evidence-based intervention to students with disabilities, and provide opportunities for content-area teachers to acquire and use instructional practices to *support* more students. CLC is based on more than 35 years of school-based research (e.g., Deshler et al. 2006; Ehren et al. 2010; Lenz et al. 2005; Schumaker and Deshler 2010) originating from the University of Kansas Institute for Research on Learning Disabilities in 1978. It provides a multi-tiered framework designed to increase students' content literacy performance within and across secondary content courses. Its teaching routines and learning strategies are robust for any content-area class that requires reading and literacy. See Schumaker and Deshler (2010) for more detail.

CLC offers five levels of increasing intensity ranging from core instructional enhancements to intensive intervention(s) outside the core curriculum. In the first levels of CLC, the emphasis is on 14 *content enhancement routines* that provide co-teachers *and* content-area teachers direct and explicit instructional tools and strategies to accomplish four broad tasks for content literacy skill acquisition:

1. Planning and organization of content so the learning process and outcomes are explicit for students;
2. Explaining texts using specific strategies for clarifying, framing, vocabulary, and analyzing complex text;
3. Teaching complex concepts for understanding and shared vocabulary for high-level discussions about text; and
4. Specific strategies for students to complete classroom work using critical thinking skills such as questioning, summarizing, planning and problem-solving, and evaluation.

Research provides ample evidence that these strategies increase content learning outcomes for secondary special education students, for below grade-level, and for average-achieving students (e.g., Bulgren 2006; Deshler et al. 2001; Swanson and Deshler 2003). For example, Bulgren et al. (2011) examined the effects of one of the 14 routines, question-exploration routine (QER) and an associated graphic organizer on grade 7 students' abilities to think about and answer complex questions. Using a counter-balanced design, moderate to very large effect sizes (0.31 to 1.6) were observed across all subgroups. Other studies showed significant increases in skills such as analyzing and evaluating a claim (Ellis and Bulgren 2009); comparing (e.g., making analogies) and contrasting conceptual information (Bulgren et al. 2000; Bulgren et al. 2002), and tracing causal reasoning and making decisions (Bulgren et al. 1998).

Explicit, Research-Based Program-Driven Intensive Basic Skills Instruction MTSS also provides another opportunity to deliver basic skill intensive intervention service delivery options using well-designed and field-tested curricula with embedded research-based instructional practices (Diamond 2004; Marchand-Martella et al. 2013; McPeak and Trygg 2007). In contrast to prevailing practices (e.g., reactive content-area tutoring, help with homework), the focus shifts to selecting and delivering effective and comprehensive *programs* similar to standard treatment protocol approach (Pericola et al. 2003). Intervention programs are reviewed reflecting a range of intensities. Based on the range of reading needs among special education students, choices are made from programs along the intensity and explicit-

Table 5 An option for how tiered reading/language arts services can be delivered

Tier	Program and focus	Staffed by	Amount of time
General education tier 1	Strong, teacher-led, comprehensive language arts program with explicit instruction in comprehending narrative and content textbooks (i.e., *Read to Achieve*) + novel study strongly biased toward nonfiction	Content-area language arts teachers	Double period or block every day
Tier 2	*Read to Achieve,* plus more explicit and targeted intervention + (e.g., *rewards*) + structured outside wide reading	Assigned content-area language arts teachers	Tier 2 delivered within the double period/block
Tier 3 and special education	*Read to Achieve* + explicit and comprehensive intervention (e.g., *REACH* or *Corrective Reading*) + structured outside wide reading	Special education personnel	3 periods

ness continuum, with *Corrective Reading* at the most intensive continuum end to *FLEX Literacy*, Language!, FUSION, with other programs like REWARDS, and Read 180 at the lesser, but still intensive end.

Once identified, these interventions are best delivered through scheduled, additional supplemental (i.e., in addition to the language arts or core content courses) periods of individualized length reflecting the severity of any specific student's performance discrepancy. The critical component is to deliver an intensive intervention as early and powerfully in secondary settings as possible, irrespective of consequences including scheduling difficulties and obtaining sufficient numbers of credit hours for graduation. The primary advantage of this option is an increased capacity to address the performance discrepancy at grade 6 and by increasing a discrepant student's basic skills, improving the odds for content-area success in each and every subsequent grade.

An illustration of how tiered services could be designed, staffed, and delivered is shown in Table 5. Among the choices for a middle school languages arts research-based core program could be *Read to Achieve* (Marchand-Martella and Martella 2010) or *Prentice Hall* (Feldman et al. 2006).

In this implementation model, tier 2 intervention to below average students would be delivered within the period, circumventing the need for students to miss other classes/courses. Intervention would not simply be reteaching, but using carefully selected intervention programs aligned with students' needs. For example, many middle school struggling reading have difficulty with multisyllabic words, vocabulary, and comprehension (Torgesen 2004). A viable solution in this instance could be REWARDS (Archer and Gleason 2001), an intervention program with strengths in these skill areas. Tier 2 intervention would be delivered by language arts teachers. Tier 3 would be targeted to students with severe performance discrepancies in an additional period, using intensive, specially designed curriculum and staffed by special education personnel.

Adopting Scientifically Based Basic Skills, IEP Goals, and Frequent Progress Monitoring

It is relatively easy to identify those middle and high school students with basic skills performance discrepancies that place them at risk or are severe enough to require intensive intervention. In theory, any good basic skills achievement test may enable schools to identify potential students who need intervention. What is historically more difficult is to increase middle and high schools' capacities to measure progress in reducing basic skill discrepancies for students who

Table 6 Recently observed IEP goals for a grade 10 student with a significant reading performance discrepancy

Goal	Criterion	Method	Frequency
XX will decode words containing long vowel syllable patterns 80 %	80 %	Documented observation	End of grading period
Frodo will decode words containing the silent syllable pattern (CVCe) 80 %	80 %	Documented observation	End of grading period
Frodo will decode words containing inflected endings (ing, ed, er, y, ly, ful) 80 %	80 %	Documented observation	End of grading period
Frodo will recognize sight words at the fourth-grade level	80 %	Documented observation	End of grading period

receive intervention. A powerful tool to increase the quality and quantity of basic skills progress monitoring is to change how special education IEP goals are written and progress is monitored. Despite decades of federal efforts to improve practices, special education IEP goals remain of poor quality, and are especially so in secondary settings (Bateman and Linden 2006; Yell and Busch 2012; Yell et al. 2009). See for example, these IEP goals in Table 6 that were written for a grade 10 student with a reading performance discrepancy.

Goals such as these examples are inadequate and inappropriate and do not form a credible basis for evaluating the effectiveness of the student's special education intervention (Shinn and Shinn 2000; Yell and Busch 2012). These goals reflect little change from practice inadequacies identified 30 years ago (Deno et al. 1984) and have direct consequences for students with disabilities regardless of MTSS.

Furthermore, there are indirect consequences for MTSS of poor IEP goals and progress monitoring. Failure to engage in well-established goal setting and progress monitoring for students *with* IEPs diminishes the capacity for high-quality tier 2 and tier 3 basic skills progress monitoring. Finally, without high-quality tier 3 basic skills progress monitoring, it is difficult, if not impossible, to use an RTI approach for determining special education eligibility for SLD using a dual discrepancy.

To increase the quality of services for students with disabilities and increase local capacity for high-quality tier 2 and tier 3 progress monitoring, best practices for writing IEP goals and progress monitoring (e.g., Fuchs and Fuchs 2008) should be followed. For reading, a single, robust IEP annual goal such as *student X will read 150 words correctly from grade 7 passages by October 12, 2014* (i.e., the date of the annual IEP meeting) meets this professional standard. Student progress would be monitored weekly comparing the student's actual rate of progress versus the expected ROI. In other words, the focus of IEP progress monitoring is the progress discrepancy. See Shinn and Shinn (2000) for more details on writing observable, measurable goals using CBM.

Teaming Strategies for Successful Secondary Implementation

Ensuring ongoing MTSS implementation success is facilitated by regular meetings of teams at least once per month to review progress of students who receive tiered interventions. At the middle school level, for tier 2 students, the likely candidates for assuming this responsibility could be either grade-level teams or departmental teams, depending on the school structure. In either case, decision-making is enhanced by including relevant administrators and related services personnel (e.g., school psychologists, counselors) who may contribute to necessarily additional intervention support services. Tier 3 decision-making should be guided by more regular (e.g., weekly, biweekly), shorter meetings of special education teams, plus administrative and related service support. At the high school level, with the exception of schools with alternative structure such as “pods,” tier 2 decision-making should be the responsibility of department teams, with appropriate support. Again, responsibility for tier 3 students rests with special education teams.

To support effective communication across building staff and years, documentation of decision-making is essential. Research supports the use of written documentation usually in the form of a checklist with precise information (e.g., who, what, when, where, and how to ensure integrity of follow through with decisions and intervention protocols) (Buffum, et al. 2012; Gansle and Noell 2007; Gawande 2009; Wickstrom et al. 1998; Windram et al. 2012). Furthermore, this documentation serves as valuable communication as students transition within and between middle and high school where increased numbers of adults and students make the channels for communicating individual and group information more complex. Examples of forms for middle and high school use may be found in Windram et al. (2012).

Adjustment of Grading Practices Study of grading procedures at the secondary level shows that clarity in expectations and a focus on demonstration of skill mastery are preferable to other grading procedures in increasing student achievement (Reeves 2008; Wormeli 2006). Some researchers have suggested the use of “mainstream consultation agreements (Tindal et al. 1987; Tindal and Germann 1991) which are agreements between a student and teacher providing operationalized expectations of student behaviors and a system for monitoring student behavior toward success in the class. For example, all classroom expectations for participation, work completion, and academic work quality can be assigned points, and the student’s accumulation of points over time may be linked to a final grade.

Finally, high schools will need to consider how to manage tier 2 and tier 3 time with credit and graduation requirements. Often, graduation requirements function as a barrier that prevents students from being able to have extra time for remedial instruction in their schedule (Windram et al. 2012). For example, many districts will allow a student to receive an academic intervention during a study hall period, but will not allow the student to earn any credit toward graduation for this time. Consequently, a plan for credit recovery is a nonnegotiable practice for high schools within MTSS. Credit recovery must not only provide additional learning time for basic academic skills but also provide study and organizational skills, problem-solving skills, prosocial interpersonal skills and behaviors (i.e., working collaboratively with others) that will prepare students for the expectations of college, career, and civic life.

From an organizational perspective, system decision-makers must intentionally prioritize the prevention effects discussed earlier in this chapter, that is intensive instructional effort to reduce and prevent performance discrepancies. Inevitably, such prioritization will create a tension in grading practices, satisfaction of course requirements, and attaining credit requirements necessary for graduation. When that tension arises, the decision-making team must make the adjustments needed to honor the priority of improving student skill proficiency and preventing sustained or worsening skill discrepancies.

Adjustment and Management of the Master Schedule The master schedule is one of the most commsonly identified challenges by secondary personnel for RTI implementation in secondary settings (National High School Center, National Center on Response to Intervention, and Center on Instruction). Often secondary schedules are rigidly bound by credit and/or “seat time” requirements that at best hinder and at worst impede students from receiving instruction appropriate to their learning level and needs, and achieving graduation.

Successful examples of how middle and high school master schedules have been adjusted to accommodate additional time for interventions may be found in Buffum et al. (2012) and Windram et al. (2012). Windram et al. (2012) suggested examining course offerings and potentially eliminating courses that do not meet minimum enrollment requirements, have high failure rates, have a history of low enrollment, or are not well-aligned to state standards. Decisions about course elimination should be considered in concert with student achievement data, graduation requirements, and school improvement plans. Eliminating less fruitful courses opens up space and personnel within the schedule to allow for the development of needed supplemental supports. Further, advisory periods and study hall periods can be repurposed for targeted instruction in content or in basic skill areas for students requiring additional support (Dixon, 2008, 2009).

Research Findings and Future Directions

The National High School Center, National Center on Response to Intervention, and Center on Instruction (2010) conducted interviews with educators from eight geographically diverse high schools self-reporting RTI practices from 1–7 years. Implementers reported that the essential features of RTI in secondary schools included: core instruction that emphasized alignment with state standards and content literacy instruction, using course failure rates (particularly for Algebra and English) as screening indicators, a clearly defined and consistent system of progress monitoring, regular use of data by teams for decision-making, an identified RTI leadership team, and targeted, sustained professional development. Further, all schools featured multi-tiered supports targeting literacy, mathematics, and/or positive-behavior interventions with an emphasis on supporting 9^{th} grade students, although no outcome data were provided. Emerging case and anecdotal data from a wide range of geographically and demographically diverse secondary schools report promising gains in student outcomes with the use of RTI in middle and high schools (Prewett et al. 2012; Stepanek and Peixotto 2009; Windram et al. 2007). For example, Windram et al. (2007) described a pilot program for rural secondary students identified as being at-risk for school failure on a group-administered achievement test and *curriculum based measures of reading*. The pilot included doubling the amount of time allocated to reading instruction, providing targeted and individualized subskill remediation for decoding, fluency, vocabulary, and comprehension, frequent progress monitoring, and regular conferencing with students to review progress and set short-term achievement goals. This intervention resulted in an average growth rate on the NWEA MAP test of more than three times the national average among students in grade 9 and more than five times their growth from the previous year. A similar pilot for mathematics conducted with eighth-grade students resulted in growth that exceeded the national average by a factor of almost six.

In two meta-analyses for students in grades 6–12, interventions targeting fluency, decoding, direct vocabulary instruction, and direct instruction on comprehension skills, yielded moderate to strong mean weighted effect sizes (e.g., 0.89 in Edmonds et al. 2006 and 0.95 in Scammaca et al. 2007). Word-level interventions showed moderate effects (0.34–0.42). Limitations noted were that the reading interventions were relatively short in duration (about 2 months), and the large effect size did not necessarily close the gap between the at-risk student's performance and the performance of not-at-risk peers.

One large-scale, randomized, year-long study of middle school students with reading difficulties applied within an RTI framework yielded more modest effect sizes (Vaughn et al. 2010), but provided fruitful ground for further research. Students from seven middle schools were assigned to treatment and control groups school-wide. All students were provided an increase in content-area vocabulary and comprehension instruction throughout the school day (tier 1). In addition, the treatment group received tier 2 (supplemental) reading interventions in medium-sized groups (10–15 students). The median effect size for increases in decoding, fluency, and comprehension skills was small at 0.16. Vaughn et al. (2010) suggested that the large sample size, length of time, and enhanced core instruction for both treatment and control groups likely influenced the effects. These findings highlight a need in RTI research to study both the overall effect of integrated systems of intervention support, and also to parcel the effects of each layer of intervention. The need to address skill gaps with older learners has caused some scholars to emphasize provision of intensive instruction immediately rather than following a process of increasing intensive supports with serial assessments (Deshler and Kovaleski 2007; Fuchs et al. 2010), which would also affect sample selection procedures and associated effect size estimates for various interventions.

Conclusion

The implementation of RTI in secondary settings shows promise for increasing student achievement; yet, research at the secondary level has lagged considerably behind elementary RTI implementation, as special challenges to effective implementation at the secondary level have been identified. This chapter provides a guideline of research-supported MTSS practices for secondary schools that is both programmatically and developmentally relevant for secondary educators and students, and creates basic academic and behavioral interventions of suitable intensity for struggling learners.

Disclosures Mark R. Shinn is a paid consultant for Pearson Assessment and consults with them regarding *aimsweb*, a commercial product that organizes and reports Curriculum-Based Measurement (CBM) data. He also receives royalties from McGraw Hill Education for his contributions *to Jamestown Reading Navigator*, an intermediate and secondary level remedial reading program. Holly S. Windram has done contracted work for FAST at the University of Minnesota. Chapter authors have no commercial interests in other programs identified in this chapter.

References

Archer, A., & Gleason, M. (2001). *Rewards: Reading excellence word attack and rate development strategies*. Longmont: Sopris West.

Barth, A. E., Stuebing, K. K., Fletcher, J. M., Cirino, P. T., Francis, D. J., & Vaughn, S. (2012). Reliability and validity of the median score when assessing the oral reading fluency of middle grade readers. *Reading Psychology, 33*(1–2), 133–161.

Bateman, B. D., & Linden, M. A. (2006). *Better IEPs: How to develop legally correct and educationally useful programs* (4th ed.). Verona: Attainment Company.

Baumann, J. F., & Kame'enui, E. J. (2004). *Vocabulary instruction: Research to practice*. New York: Guilford Press.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.

Buffum, A., Mattos, M., & Weber, C. (2012). *Simplifying response to intervention four essential guiding principles*. Bloomington: Solution Tree.

Bulgren, J. A. (2006). Integrated content enhancement routines: Responding to the needs of adolescents with disabilities in rigorous inclusive secondary content classes. *Teaching Exceptional Children, 38,* 25–36.

Bulgren, J. A., Deshler, D. D., & Schumaker, J. B. (1998). Reasoning strategies and teaching routines for use in mainstream content classrooms. Final research report submitted to the U.S. Department of Education, Special Education Services.

Bulgren, J. A., Deshler, D. D., Schumaker, J. B., & Lenz, B. K. (2000). The use and effectiveness of analogical instruction in diverse secondary content classrooms. *Journal of Educational Psychology, 92,* 426–441.

Bulgren, J. A., Lenz, B. K., Schumaker, J. B., Deshler, D. D., & Marquis, J. G. (2002). The use and effectiveness of a comparison routine in diverse secondary content classrooms. *Journal of Educational Psychology, 94*(2), 356.

Bulgren, J. A., Marquis, J. G., Lenz, B. K., Deshler, D. D., & Schumaker, J. B. (2011). The effectiveness of a question-exploration routine for enhancing the content learning of secondary students. *Journal of Educational Psychology, 103,* 578–593.

Burns, M. K., Sarlo, R., & Pettersson, H. (2013). Response to intervention for literacy in secondary schools. http://rtinetwork.org/learn/rti-in-secondary-schools/rti-literacy-secondary-schools.

Carnine, D., Miller, S., Bean, R. M., & Zigmond, N. (1994). Social studies: Educational tools for diverse learners. *School Psychology Review, 23,* 428–441.

Carnine, D. W., Crawford, D. B., & Harniss, M. K. (2007). Effective strategies for teaching social studies. In M. D. Coyne, E. J. Kame'enui, & D. W. Carnine (Eds.), *Effective teaching strategies that accommodate diverse learners* (3rd ed., pp. 139–158). Upper Saddle River: Pearson, Merrill Prentice Hall.

Conderman, G., & Petersen, T. (2005). Promoting positive special education practices. *NASSP Bulletin, 89,* 90–98.

Conderman, G., & Petersen, T. (2007). Avoid the tutoring trap. *Intervention in School & Clinic, 42,* 234–238.

Cook, C. R., Sprague, J., Browning Wright, D., & Sadler, C. (2008). *RTI and behavior answer book on RTI*. Horsham: LRP Publications.

Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2007a). Effective strategies for teaching writing. In M. D. Coyne, E. J. Kame'enui, & D. W. Carnine (Eds.), *Effective teaching strategies that accommodate diverse learners* (3rd ed.). Upper Saddle River: Pearson: Merrill Prentice Hall.

Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2007b). *Effective teaching strategies that accommodate diverse learners* (3rd ed.). Upper Saddle River: Pearson: Merrill Prentice Hall.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Deno, S. L. (2012). Problem-solving assessment. In R. Brown-Chidsey & K. Andren (Eds.), *Assessment for intervention: A problem-solving approach* (2nd ed., pp. 10–36). New York: Guilford.

Deno, S. L., Mirkin, P., & Wesson, C. (1984). How to write effective data-based IEPs. *TEACHING Exceptional Children, 16,* 99–104.

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading, 15,* 109–135.

Deshler, D. D., & Cornett, J. (2012). Leading to improve teacher effectiveness: Implications for practice, reform, research, and policy. In J. B. Crockett, B. S. Billingsley, & M. L. Boscardin (Eds.), *Handbook of leadership and administration for special education*. New York: Taylor & Francis.

Deshler, D. D., & Kovaleski, J. F. (2007). Secondary applications of RtI: A guided discussion. Presentation at the response to intervention summit, Arlington, Virginia. Retrieved 15 June 2008.

Deshler, D. D., & Schumaker, J. B. (2005). *Teaching adolescents with disabilities: Accessing the general curriculum* (2nd ed.). Thousand Oaks: Corwin.

Deshler, D. D., Ellis, E., & Lenz, B. K. (1996). *Teaching adolescents with learning disabilities* (2nd ed.). Denver: Love.

Deshler, D. D., Schumaker, J. B., Lenz, B. K., Bulgren, J. A., Hock, M. F., Knight, J., & Ehren, B. (2001). Ensuring content-area learning by secondary students with learning disabilities. *Learning Disabilities Research & Practice, 16*(2), 96–108.

Deshler, D. D., Hock, M. F., & Catts, H. W. (2006). Enhancing outcomes for struggling adolescent readers. *Perspectives (Gerontological Nursing Association (Canada)), 32,* 21–25.

Diamond, L. (2004). *Implementing and sustaining a middle and high school reading and intervention program*. Berkeley: Consortium on Reading Excellence.

Dixon, R. D. (2008). Getting the school ready for the adolescents: Designing RtI Practices at the secondary level. Presentation at the annual meeting of the Minnesota Association of School Psychologists, St. Paul, MN.

Dixon, R. D. (2009). Getting the School Ready for the Adolescents: Designing RtI Practices at the Secondary Level. Presentation at the Wisconsin RtI Summit, Green Bay, WI.

Duffy, H. (2007). *Meeting the needs of significantly struggling learners in high school: A look at approaches to tiered intervention*. Washington, DC: National High School Center, American Institute for Research (AIR).

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving student learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58.

Ehren, B. J., Deshler, D. D., & Graner, P. S. (2010). Using the content literacy continuum as a framework for implementing RtI in secondary schools. *Theory into Practice, 49,* 315–322.

Ellis, J. D., & Bulgren, J. (2009). Improving teaching of scientific argumentation skills. Paper presented at the annual meeting of the Association for Science Teacher Educators, Hartford.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction*. New York: Irvington Publishers.

Engelmann, S., Hanner, S., & Johnson, G. (1999). *Corrective reading series*. Columbus: McGraw Hill Education.

Espin, C., & Campbell, H. M. (2012). They're getting older…but are they getting better? The influence of curriculum-based measurement on programming for secondary-school students with learning disabilities. In C. A. Espin, K. McMaster, S. Rose, & M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 149–164). Minneapolis: University of Minnesota Press.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62,* 497–514.

Espin, C., & Tindal, G. A. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 214–253). New York: Guilford.

Espin, C. A., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress measurement system in reading for middle-school student: Monitoring progress towards meeting high stakes standards. *Learning Disabilities Research and Practice, 25,* 60–75.

Foegen, A. (2012). Big ideas and core values: The influence of Stanley Deno's work on secondary progress monitoring. In C. A. Espin, K. McMaster, S. Rose, &

M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 139–148). Minneapolis: University of Minnesota Press.

Frey, A. J., Lingo, A., & Nelson, C. M. (2010). Implementing positive behavior support in elementary schools. In M. R. Shiinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 397–433). Bethesda: National Association of School Psychologists.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488–500.

Fuchs, L. S., & Fuchs, D. (2008). Best practices in progress monitoring reading and mathematics at the elementary level. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 2147–2164). Bethesda: National Association of School Psychologists.

Fuchs, L. S., & Vaughn, S. R. (2005). Response to Intervention as a framework for the identification of learning disabilities. Forum for trainers of school psychologists, spring, 12–19.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Rethinking response to intervention at middle and high school. *School Psychology Review, 39,* 1.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239–256.

Gansle, K. A., & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. Jimerson, M. Burns, & A. VanDerHeyden (Eds.), *Handbook of response to intervention the science and practice of assessment and intervention* (pp. 244–251). New York: Springer.

Gawande, A. (2009). *The checklist manifesto*. New York: Henry Holt Publishers.

Germann, G. (2010). Thinking of yellow brick roards, emerald cities, and wizards. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. xiii–xxxv). Bethesda: National Association of School Psychologists.

Germann, G. (2012). Implementing data-based program modification: Big ideas. In C. A. Espin, K. McMaster, S. Rose, & M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 79–87). Minneapolis: University of Minnesota Press.

Gersten, R., Beckman, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools* (pp. 1–104). Washington, DC: US Department of Education.

Gresham, F., Reschly, D., & Shinn, M. R. (2010). RTI as a driving force in educational improvement: Historical legal, research, and practice perspectives. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 47–77). Bethesda: National Association of School Psychologists.

Grossen, B., Romance, N. R., & Vitale, M. R. (1994). Science: Educational tools for diverse learners. *School Psychology Review, 23,* 442–463.

Harris, K. R., Graham, S., Mason, L. H., & Friedlander, B. (2008). *Powerful writing strategies for all students*. Baltimore: Paul H. Brookes Publishing.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Houghton Mifflin Harcourt Riverside Assessment. (2013). *easyCBM*. Rolling Meadows: Houghton Mifflin Riverside Assessment.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*(5), 421–432.

Jenkins, J. R., & Fuchs, L. S. (2012). Curriculum-based measurement: The paradigm, history, and legacy. In C. A. Espin, K. McMaster, S. Rose, & M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 7–23). Minneapolis: University of Minnesota Press.

Kame'enui, E. J., & Baumann, J. F. (Eds.). (2004). *Vocabulary instruction: Research to practice*. New York: Guilford Press.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. K. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Educational Sciences, U.S. Department of Education.

Lenz, B. K., Deshler, D. D., & Kissam, B. R. (2003). *Teaching content to all: Evidence-based inclusive practices in middle and secondary schools*. Boston: Allyn & Bacon.

Lenz, B. K., Ehren, B. J., & Deshler, D. D. (2005). The content literacy continuum: A school-reform framework for improving adolescent literacy for all students. *TEACHING Exceptional Children, 37*(6), 60.

Marchand-Martella, N. E., & Martella, R. C. (2010). *Read to achieve: Comprehending content-area text and comprehending narrative text*. Chicago: McGraw Hill Education.

Marchand-Martella, N. E., Martella, R. C., Modderman, S. L., Petersen, H., & Pan, S. (2013). Key areas of effective adolescent literacy programs. *Education and Treatment of Children, 36,* 161–184.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52,* 266–276.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria: Association for Supervision and Curriculum Development.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing. *The Journal of Special Education, 41*(2), 68–84.

McPeak, L., & Trygg, L. (2007). *The secondary literacy instruction and intervention guide: Helping school districts transform into systems that produce life-changing results for all children*. Mill Valley: Stupski Foundation.

National Center for Education Statistics. (2013). *The nation's report card: a first look: 2013 mathematics and* reading (NCES 2014–451). Washington, DC: Institute for Education Sciences, US Department of Education.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2012). Common Core State Standards Initiative. http://www.corestandards.org.

National Mathematics Advisory Panel. (2008). *Foundations for success: National mathematics advisory panel final report*. Washington, DC: US Department of Education.

Northwest Evaluation Center. (2014). Measures of academic progress. Portland: NWEA.

Pericola Case, L., Speece, D. L., & Eddy Molloy, D. (2003). The validity of response-to-instruction paradigm to identify reading disabilities: A longitudinal analysis of individual differences and context factors. *School Psychology Review, 32,* 557–582.

Phillips, M. Dissertation. (2002). An examination of the instructional needs of secondary students with learning disabilities and the reading interventions they receive. Unpublished doctoral dissertation, Eugene. University of Oregon.

Prewett, S., Mellard, D. F., Deshler, D. D., Allen, J., Alexander, R., & Stern, A. (2012). Response to intervention in middle schools: Practices and outcomes. *Learning Disabilities Research & Practice, 27,* 136–147.

Reeves, D. (2008). Effective grading practices. *Educational Leadership, 65,*(5), 85.

Renaissance Learning. (2012). *STAR reading technical manual*. Wisconsin Rapids: Renaissance Learning.

Rose, T. E., & Zirkel, P. A. (2007). Orton-Gillingham methodology for students with reading disabilities: 30 years of case law. *The Journal of Special Education, 41*(3), 171–185.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth: RMC Research Corporation, Center on Instruction.

Scholastic Inc. (2001). *Scholastic reading inventory*. New York: Scholastic Inc.

Schumaker, J. B., & Deshler, D. D. (2010). Using a tiered intervention model in secondary schools to improve academic outcomes in subject-area courses. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior in a 3-tier model including RTI* (pp. 609–632). Bethesda: National Association of School Psychologists.

Schuta, T., Mauricio, D., & Comerford, S. (2012). Significant steps forward. *Principal Leadership* (November), 32–35.

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.

Shinn, M. R. (Ed.). (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.

Shinn, M. R. (2008). RTI at the secondary level. In S. A. Fernley, S. D. LaRue, & J. W. Norlin (Eds.), *What do I do when… The answer book on RTI*. Horsham: LRP Publications.

Shinn, M. R. (2010). Building a scientifically based data system for progress monitoring and universal screening across three tiers including RTI using curriculum-based measurement. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 259–293). Bethesda: National Association of School Psychologists.

Shinn, M. R., & Shinn, M. M. (2000). Writing and evaluating IEP Goals and making appropriate revisions to ensure participation and progress in general curriculum. In C. F. Telzrow & M. Tankersley (Eds.), *IDEA Amendments of 1997: Practice guidelines for school-based teams* (pp. 351–381). Bethesda: National Association of School Psychologists.

Spectrum K12 School Solutions. (2011). Response to Intervention Adoption Survey 2011: Global Scholar/Spectrum K12.

Sprague, J. R., & Walker, H. M. (2010). Building safe and healthy schools to promote school success: Critical issues, current challenges, and promising approaches. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 225–258). Bethesda: National Association of School Psychologists.

Sprick, R. S. (2006). *START on Time! Safe transitions and reduced tardies*. Eugene: Safe and Civil Schools/Pacific Northwest Publishing.

Sprick, R. S., & Borgmeier, C. (2010). Prevention and management of behavior in three tiers at the secondary level. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (pp. 435–468). Bethesda: National Association of School Psychologists.

Sprick, R. S., & Garrison, M. (2008). *Interventions: Evidence-based behavioral strategies for individual students* (2nd ed.). Eugene: Pacific Northwest Publishing, Inc.

Stahl, S., & Kapinus, B. (2001). *Word power: What every educator needs to know about teaching vocabulary*. Washington, DC: National Education Association.

Stepanek, J., & Peixotto, K. (2009). Models of response to intervention in the northwest region states. Issues & Answers Report, REL, (079).

Swanson, H. L., & Deshler, D. (2003). Instructing adolescents with learning disabilities converting a meta-analysis to practice. *Journal of Learning Disabilities, 36*(2), 124–135.

Thompson, L. J. (2006). *Preparing a broken foundation: Solutions to help middle school students struggle with math (Vol. Fall)*. Dallas: Voyager

Tindal, G. A., & Germann, G. (1991). Mainstream consultation agreements in secondary schools. In G. Stoner, M. R. Shinn, & H. M. Walker (Eds.), Interventions for achievement and behavior problems (pp. 495–518.). Bethesda: NASP.

Tindal, G. A., Wesson, C., Deno, S. L., Germann, G., & Mirkin, P. (1985). The pine county model for special education delivery: A data-based system. In T. Kratochwill (Ed.), *Advances in school psychology* (Vol. IV, pp. 223–250). Hillsdale: Lawrence Erlbaum Assoc.

Tindal, G. A., Shinn, M. R., Walz, L., & Germann, G. (1987). Mainstream consultation in secondary settings: An evaluation of the Pine County Model. *The Journal of Special Education, 21,* 94–106.

Tolar, T., Barth, A. E., Francis, D. F., Fletcher, J. M., Stuebing, K. K., & Vaughn, S. (2012). Psychometric properties of maze tasks in middle school students. *Assessment for Effective Intervention, 37*(3), 131–146.

Torgesen, J. K. (2004). Lessons learned from the last 20 years of research in the interventions for students who experience difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 225–229). Baltimore: Brookes Publishing.

Torgesen, J. K., Nettle, S., Howard, P., & Winterbottom, R. (2005). Brief report of a study to investigate the relationship between several brief measures of reading fluency and performance on the florida comprehensive assessment test-reading in 4th, 6th, 8th, and 10th grades. FCRR technical report 6. 2005. http://www.fcrr.org/TechnicalReports/Progress_monitoring_report.pdf.

Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S., Wexler, J., Francis, D. J., Rivera, M. O., & Lesaux, N. (2007a). *Academic literacy instruction for adolescents: A guidance document from the center on instruction*. Portsmouth: RMC Research Corporation Center on Instruction.

Torgesen, J. K., Houston, D., & Rissman, L. (2007b). *Improving literacy instruction in middle and high schools: A guide for principals*. Portsmouth: RMC Research Corporation, Center for Instruction.

Troia, G. A., & Olinghouse, N. G. (2013). The common core state standards and evidence-based educational practices: The case of writing. *School Psychology Review, 42,* 343–357.

Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. D., et al. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 30,* 3–21.

Walker, H. M. (2004). Commentary: Use of evidence-based interventions in schools: Where we've been, where we are, and where we need to go. *School Psychology Review, 33,* 398–407.

Walker, H. M., & Shinn, M. R. (2010). Systemic, evidence-based approaches for promoting positive student outcomes within an RTI framework: Moving from efficacy to effectiveness. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 1–26). Bethesda: National Association of School Psychologists.

Wallace, T., Espin, C. A., McMaster, K., Deno, S. L., & Foegen, A. (2007). CBM progress monitoring within a standards-based system: Introduction to the special series. *The Journal of Special Education, 41,* 66–67.

Ward, B., & Gersten, R. (2013). A randomized evaluation of Safe and Civil Schools model for positive behavioral interventions and supports at elementary schools in a large urban district. *School Psychology Review, 42,* 317–333.

Wickstrom, K. F., Jones, K. M., La Fleur, L. H., & Mortenson, B. P. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13,* 141–154.

Windram, H., Bollman, K., & Johnson, S. (2012). *How RtI works in secondary schools building a framework for success*. Bloomington: Solution Tree.

Windram, H., Scierka, B., & Silberglitt, B. (2007). Response to intervention at the secondary level: A description of two districts' models of implementation. *Communiqué, 35,* 43–45.

Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education, 34,* 14–27.

Yell, M. L., & Busch, T. W. (2012). Using curriculum-based measurement to develop educationally meaningful and legally sound individualized educational programs. In C. A. Espin, K. McMaster, S. Rose, & M. Wayman (Eds.), *A measure of success: The influence of curriculum-based measurement on education* (pp. 79–87). Minneapolis: University of Minnesota Press.

Yell, M. L., Ryan, J. B., Rozalski, M. E., & Katsiyannis, A. (2009). The US Supreme Court and special education: 2005 to 2007. *TEACHING Exceptional Children, Jan/Feb,* 68–75.

Zigmond, N. (1990). Rethinking secondary school programs for students with learning disabilities. *Focus on Exceptional Children, 23*(1), 1–24.

Advances in Multi-Tiered Systems of Support for Prekindergarten Children: Lessons Learned from 5 Years of Research and Development from the Center for Response to Intervention in Early Childhood

Judith J. Carta, Charles R. Greenwood, Howard Goldstein, Scott R. McConnell, Ruth Kaminski, Tracy A. Bradfield, Alisha Wackerle-Hollman, Maura Linas, Gabriela Guerrero, Elizabeth Kelley and Jane Atwater

The past 5 years have seen a growing awareness that response to intervention (RTI) models or approaches that incorporate multi-tiered systems of support (MTSS) hold great promise for children in the preschool years (Buysse and Peisner-Feinberg 2013; Carta and Greenwood 2013). RTI models used in prekindergarten programs and other early childhood settings have the potential to strengthen the link between early childhood education and elementary school (Pierce and Bruns 2013), reduce incidence of children who require special education in later grades (Burns et al. 2005), and address the school readiness needs of children with disabilities and those at risk for developmental delays (Fletcher and Vaughn 2009). Research related to these models is advancing this growing movement to promote the academic and social outcomes of young struggling learners (Greenwood et al. 2012).

Although some early education programs in the USA have incorporated RTI approaches into their models for a number of years (e.g., Chandler et al. 2008), the implementation of tiered models in preschool contexts presents some unique challenges (Greenwood et al. 2011a). To understand these challenges, it is important to understand the many ways in which early education programs are different from K–12 settings. At the forefront of these differences is the wide variability of educational programs serving young children. At this point in the USA, there is no universal prekindergarten program and as a result, young

Author note: Preparation of this chapter was supported by Cooperative Agreement #R324C080011 from the Institute of Education Sciences, U.S. Department of Education. We acknowledge the contribution of our CRTIEC colleagues Annie Hommel, Naomi Schneider, and a host of dedicated research assistants at University of Kansas, University of Minnesota, the Ohio State University, and the Dynamic Measurement Group.

Scott McConnell, Tracy Bradfield, and Alisha Wackerle-Hollman have developed assessment tools and related resources known as *Individual Growth & Development Indicators* and *Get it, Got it, Go!* This intellectual property is subject of technology commercialization by the University of Minnesota, and portions have been licensed to Early Learning Labs, Inc. Scott McConnell has equity interest in Early Learning Labs, Inc., a company which may commercially benefit from the results of this research. The University of Minnesota also has equity and royalty interests in ELL which, in turn, may benefit McConnell, Bradfield, and Wackerle-Hollman. These relationships have been reviewed and are being managed by the University of Minnesota in accordance with its conflict of interest policies.

J. J. Carta (✉) · C. R. Greenwood
University of Kansas, Kansas City, KS, USA
e-mail: carta@ku.edu

H. Goldstein
University of South Florida, Tampa, FL, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_33

children are served in a variety of settings. Indeed, even identifying the percentage of children enrolled in preschool is not a straightforward proposition in that a number of educational program types are available for preschool-aged children. Defining "preschool" is variable. Does "preschool" include only state-funded prekindergarten or does it include Head Start and privately funded early education programs? Does "preschool" also include child care? As of 2012, approximately 31 % of 4-year-olds across the nation were enrolled in either general or special education state-funded prekindergarten programs. This percentage grew to 41 % when Head Start was included (Barnett et al. 2012). This figure increases to approximately 57 % when enrollment in child care is considered (Pew Charitable Trusts 2004). A further challenge in describing preschool settings is the variability across child care environments. Child care programs can vary from highly resourced, high-quality early education programs to family-run childcare programs offering little in the way of a school readiness curriculum. Therefore, designing MTSS that can fit the variety of settings in which preschool-aged children are educated is a daunting task.

This variability of program types creates a wide-ranging preschool landscape with considerable variation across a number of different important dimensions that are relatively standard in K–12 settings. For example, teaching credentials or levels of training expected of classroom staff differ widely across early education settings, with state-funded pre-K programs typically requiring at least BA degrees, Head Start programs moving toward but not yet requiring BA-level teachers, and community-based child care often requiring no basic teaching credential. This difference in levels of teaching credentials is important because many studies have found a high correspondence between teacher education and training and higher quality programs (Burchinal et al. 2002; National Institute of Child Health and Human Development Early Child Care Research Network 2000, 2002; Vandell 2004). The availability of various teaching staff also varies considerably across settings. School district-affiliated prekindergarten programs often have a lead teacher with a minimum of a bachelor's degree and a paraprofessional, or two BA-level lead teachers with certification. They also may have access to a full complement of ancillary staff (e.g., school psychologists, speech and language pathologists). Head Start programs typically have more limited staffing patterns, and other early education classrooms may operate with only one teacher. Therefore, in terms of implementation of tiered models, the issue of who will conduct higher tiered instructional practices or who will carry out the assessments needed to identify children or monitor progress is an important concern.

Another major challenge to the implementation of tier models in early childhood settings is the variation and relatively weak capacity in curricular supports available for prekindergarten programs. While research has both expanded the knowledge of skills that children develop during the preschool years that are sensitive to intervention (e.g., National Early Literacy Panel 2009), controlled evaluations of the effects of contemporary curricula generally indicate unreliable or weak effects on individual children's development (Preschool Curriculum Evaluation Research Consortium 2008). While researchers and policy-makers continue to strive for more systematic, and more substantial, effects due to curriculum and instruction, substantial development in this area is still needed.

Finally, across the variety of early childhood programs, there is no standardization of the hours of operation. Some programs may serve children for half days and may only operate a few days a week. When programs have more limited hours overall, this restricts the time available for incor-

S. R. McConnell · T. A. Bradfield · A. Wackerle-Hollman
University of Minnesota, Minneapolis, MN, USA

R. Kaminski
Dynamic Measurement Group, Eugene, OR, USA

M. Linas
University of Kansas, Minneapolis, MN, USA

G. Guerrero · J. Atwater
University of Kansas, Kansas City, KS, USA

E. Kelley
University of MIssouri, Minneapolis, MN, USA

porating higher tiers of instruction for children identified as needing more intensity or greater instructional support. Therefore, accommodating instructional activities that provide higher tiers of support within a time-crunched program schedule can be a considerable challenge.

These features of prekindergarten settings were just some of the issues the Center for RTI in Early Childhood (CRTIEC) considered when it set out to design and validate some of the most critical features that would be essential to programs seeking to develop multi-tiered approaches in the area of early literacy and language. While recognizing that preschool children have needs for enhanced instruction across many domains, CRTIEC selected early literacy and language knowing the central importance of these areas to lifelong independence, productivity, and participation (National Early Literacy Panel (NELP), Shanahan and Lonigan). CRTIEC therefore set out to develop and validate the following:

1. An identification, progress-monitoring, and decision-making approach that would help specify the most appropriate instructional tier of support for children in early literacy and language.
2. For children not progressing in tier 1, a set of tier 2 interventions provided enhanced instruction with increased opportunities to respond or practice language or literacy skills.
3. For children not progressing in tier 2, a set of tier 3 interventions focused on a restricted set of skills known to be most critical for later school success and taught through the use of explicit instruction, carefully designed examples and instructional materials, sufficient practice of new skills to achieve fluency, and systematic review of previously learned skills.

A fundamental assumption of this work was that a tiered system of support in early literacy and language could only be successful in programs that were implementing high-quality tier 1 instruction focusing on the four instructional domains known to be most highly related to later success in reading: oral language, phonological awareness, letter/sound knowledge, and print awareness (NELP; Shanahan and Lonigan). Therefore, a fourth goal of CRTIEC was carrying out a descriptive study of the quality of early literacy and language instruction carried out in tier 1 across a diverse array of early education programs. The purpose of this chapter is to describe briefly the research conducted in each of the four areas above, highlight some of the lessons learned in carrying out work in each area, discuss the implications of the findings for practice, policy, and research, and project some of the future research that is needed in each area.

Investigating the Foundation of RTI: What Was the Quality of Tier 1 Instruction?

Overview of CRTIEC Research

Prevention-oriented approaches to preschool instruction like RTI offer a means of increasing the intensity of instruction children receive compared to the "business as usual, one size fits all instruction." Therefore, a multisite investigation to answer the questions was conducted: "What exactly is 'business as usual' in preschool Tier 1 early literacy and language instruction?" and "How are children across these programs responding to the given level of literacy instruction?".

Sixty-five classrooms from 23 programs/districts in four states serving 659 children in their year prior to kindergarten were selected (Greenwood et al. 2012). Because instructional interactions between teachers and students are a critical quality factor in promoting educational experiences and child outcomes, direct, observational measurement of these interactions in the measurement design along with formative and summative measures of children's language and literacy outcomes at the beginning, middle, and end of the year prior to kindergarten were included.

Lessons Learned About Tier 1 Instruction

There were sizable percentages of children (30 %) in the entire sample at the start of the year with weak skills or very weak skills based on the *Get Ready To Read Screener-R* (Lonigan and Wilson 2008). The highest proportions of children with weak early literacy and language skills were in programs with poverty-based eligibility requirements such as state-funded pre-K programs, Title 1, and Head Start compared to tuition-based programs. Between 30 and 40 % of children in income-based programs demonstrated weak or very weak early literacy/language skills compared to only 10 % of children who were in tuition-based programs. On average, children in the entire sample made positive gains in language and literacy skills on both formative and standardized, norm-referenced test measures. In nearly all cases, however, gaps in skill levels remained relatively unchanged over the course of the year between groups of children identified in the Fall with weak (tier 2) or very weak (tier 3) skill levels compared to children who were not identified as at-risk in the Fall. This finding demonstrated that the children with weak skills who were receiving only tier 1 instruction were not closing the gap in their early literacy knowledge and would probably enter kindergarten still significantly below not-at-risk peers if supplemental instructional support was not provided.

The study also provided a window into the amount of literacy instruction that occurred within the 65 classrooms. Observations were conducted across a range of classroom activities in which literacy was likely to be embedded (e.g., dramatic play, centers, art, story reading, nature, music). Observations revealed that in these programs that had language and literacy outcome goals, literacy instructional quality was weak. For example, teachers were found to be focusing on literacy during only 15 % of observed time and children's level of literacy engagement was variable and relatively infrequent ($M=22.8$ %, SD = 18.0). Children's literacy engagement was strongly correlated to the amount of time teachers spent carrying out literacy activities ($r=0.71$). The mean level of general instructional quality as measured by the instructional support subscale on the *Classroom Assessment Scoring System* (*CLASS*; Pianta et al. 2008) also was low ($M=2.5$, SD = 0.76).

Implications for Practice, Policy, and Research

Considering relatively conservative standards for identifying risk and the complementary assumptions that early and preventive intervention is most effective, and that numbers of children with low academic performance only increase over the grades, the authors conclude that sizable numbers of preschool children are not on a path toward being ready for the early literacy and language expectations of kindergarten. Proactive, prevention-oriented approaches to preschool instruction like MTSS offer a means of differentiating instruction to meet the needs of individual children which represents an improvement over the "one size fits all" approach to preschool instruction that is more commonly used in most preschool settings. In this study, high proportions of children (30–40 %) were identified who would be appropriate candidates for more intensive, differentiated language and literacy instruction. The study provided evidence showing that many of these children who began their pre-k year at significant literacy risk demonstrated no progress in closing the gap in literacy skills when they were provided only with tier 1 instruction. The conclusion was that if programs seek to make progress in closing the gap and getting more children ready for kindergarten, more proactive steps are needed in implementing enhanced instruction to children showing the first signs of language and literacy delays. If schools have a goal for increasing literacy benchmarks by the end of kindergarten, investing in early intervention programs that reduce the number of children entering kindergarten with literacy delays is a wise policy.

High-quality tier 1 instruction is one of the most critical components of an effective prevention-oriented multi-tiered approach to supporting optimal growth in early literacy. In addition to the use of evidence-based curricula, teachers need help in implementing high-quality classroom instruction. A key to improvement is the use of performance-based coaching using data and feedback (McCollum et al. in press). The next generation of research on tier 1 must focus on professional development approaches that are linked to observable changes in teacher behavior and child performance. For example, performance-based coaching can focus generally on increasing teachers' literacy focus, defined as teacher talk about language and literacy topics in order to boost student literacy engagement and language and literacy outcomes. Specifically, indicators of teacher literacy focus that we have used as coaching feedback from the CIRCLE direct classroom observational measure have included: phonological awareness, alphabet and print awareness, comprehension (story and other), vocabulary, reading, and literacy involvement (Atwater et al. 2009; Greenwood et al. 2012). These teacher behaviors are aligned with what is known about important instructional targets in preschool (Gutiérrez et al. 2010; Shanahan and Lonigan), and a number of evidence-based practices teachers can readily use to promote these oral language, alphabet knowledge, and phonological awareness targets among others (i.e., writing). Examples of literacy promoting practices teachers can use to engage student literacy interactions are: *choral reading* (Ezell and Justice 2005; Justice and Kaderavek 2004) and *identify, describe, examine, ask, say* (*IDEAS*; Marulis and Neuman 2010; McGinty et al. 2012).

Tier 2 Intervention in Oral Language and Early Literacy

Research Overview

One of the major goals of the CRTIEC project was to develop tier 2 curricula that addressed areas that would contribute significantly to school readiness and potentially help prevent reading, language, and other developmental disabilities. To have the greatest impact on children's ability to read and write, four areas were targeted: vocabulary, comprehension, phonological awareness, and alphabet knowledge. As curriculum development progressed, vocabulary and comprehension instruction were combined into one curriculum (*My Story Friends*) and phonological awareness and alphabet knowledge instruction were combined into another curriculum (*PA Pathways to Literacy*).

Curriculum Development and Evaluation Curriculum development was informed by a series of studies conducted over 4 years. The general progression was from single case experimental design studies with replications across sites to randomized group designs to cluster randomized design studies. Research staff was responsible for implementing the interventions during the initial series of efficacy evaluations and refinement studies (Spencer et al. 2012). However, in later effectiveness studies, trained teachers and aides implemented the curricula (www.crtiec.org).

The tier 2 language curriculum, *My Story Friends,* was designed to be a supplemental, center-based program with feasibility of high fidelity implementation in classroom settings by any educational staff. Vocabulary and comprehension skills were targeted for intervention because they are important predictors of later academic outcomes, in particular reading comprehension (Dickinson and Tabors 2001; Duncan et al. 2007). In its current form, the *My Story Friends* program includes 26 books (divided into two series *Jungle Friends* and *Forest Friends*) with embedded lessons and accompanying audio files. Each book includes embedded instruction for two challenging vocabulary words (e.g., *brave, thrilled*), one or two basic concepts (e.g., *high, low*), and three questions about the story (e.g., *Why did the friends help Ellie?*). The program is administered in small group "listening centers" in which groups of three to four children listen under headphones to prerecorded books and lessons. An adult facilitator is present at the listening center and assists children in participating (e.g., staying on the right page in their book) and offers encouragement, but does not provide any additional instruction. Children are asked to listen to the story three times during a single session that requires approximately 15 minutes and teachers are encouraged to provide this opportunity to children during center time each day.

The tier 2 early literacy curriculum (*PA pathway to literacy*) focuses on teaching phonological awareness and alphabet knowledge skills to children who are falling behind their peers in their prekindergarten year. As in the *My Story Friends* curriculum, the *PA Pathways to Literacy* curriculum is delivered in small-group settings. Children participate in groups of three and an adult serves as the instructor. In its current form, the *Paths to Literacy* curriculum includes a sequence of 12 phonological awareness activities, with three lessons available for each task if needed. For example, the children progress from blending and segmenting compound words to two-syllable words. As the curriculum progresses, the focus shifts to the phoneme level and children progress from identifying first sounds in segmented one-syllable words to identifying first sounds in multisyllabic words. These activities are scripted so that teachers can present them within a game context. In contrast to *My Story Friends,* this curriculum is not prerecorded and the instructor provides feedback to the children contingent on the accuracy of their collective performance and provides correction individually as needed.

Lessons Learned About Tier 2 Language and Early Literacy Intervention

For the tier 2 language curriculum, efficacy studies have produced promising results (Greenwood et al. submitted; Kelley et al. 2014; Spencer et al. 2012). Research staff implemented the intervention in the early efficacy studies and these findings informed the design of manuals, teacher materials, and procedures for implementation in classroom settings by classroom teaching staff. Children who participated in the *My Story Friends* program demonstrated gains in vocabulary on curriculum-based measures. Children learned an average of 10 new challenging words out of a possible of 18 (range = 7–15) as measured by monthly unit vocabulary tests. In the group design studies, effect sizes relative to a business-as-usual control group and a storybook comparison without embedded lessons were large (d ranged from 1.00 to 2.01 across units). Results for the comprehension domain were more modest with effect sizes ranging from.15 to 1.06.

When using these materials to train teaching staff, it was found that *My Story Friends* again was shown to be efficacious in classroom settings. In the cluster randomized trial, overall fidelity of implementation was very high as anticipated given the automated nature of the vocabulary intervention. One third of the intervention sessions were videotaped and tapes were reviewed by a trained fidelity checker using an implementation checklist. Fidelity of implementation in the first *Forest Friends* series was 98.9 % and was 100 % in the *Jungle Friends* series. Classrooms completed the 26 books of intervention, although the intended dosage (three listens per book) was not achieved for all books in all classrooms. A majority of teachers rated the program positively and indicated that they believed children benefited from the program. However, some teachers indicated that it was challenging to implement the program within the context of their classroom.

The tier 2 *PA Pathways to Literacy* curriculum focused on teaching phonological awareness and alphabet knowledge skills to children who were falling behind their peers in their prekindergarten year. In its current form, the *PA Pathways to Literacy* curriculum is a scripted intervention that teaching staff implemented with three children in learning centers. The original conceptualization of this intervention, based on the research of Ziolkowski and Goldstein (2008), embedded literacy lessons within prerecorded storybooks (similar to the vocabulary lessons that were embedded in the *My Story Friends* curriculum). However, this story-embedded approach yielded inconsistent results and the intervention was modified. Procedures for providing contingent feedback and correction were incorporated to facilitate learning among children who were not benefiting from general classroom instruction on phonological awareness skills. The curriculum was revised into a game format without the storybook context and consistent improvements in children's abilities to identify the first sounds in words were found (Noe et al. 2013). In fact, by the end of the intervention, eight of nine participants performed above the benchmark for kindergartners on the *DIBELS First Sound Fluency* measure (Good et al. 2013).

Implications for Practice, Policy, and Research

The tier 2 work carried out by CRTIEC responds to the need for oral language and early literacy interventions that can produce strong effects when implemented by educational staff in preschool classrooms. The Center's development work followed an iterative process to design, implement, test, and refine intervention procedures that could work in typical early childhood education settings. Findings suggested that intentional instruction (e.g., scripted intervention procedures and protocols) that emphasized multiple opportunities to respond with individualized and immediate corrective feedback are necessary features of tier 2 preschool literacy programs.

While the goal of the tier 2 development process was the design of curricula that would be feasible and usable in a variety of early childhood settings, there may be requisite conditions in prekindergarten programs that set the stage for adoption of these curricula. First, programs and teaching staff need to be committed philosophically and pedagogically to explicit instruction of language and literacy skills. Second, programs need to provide structure that accommodates small-group learning centers for at least 15 minutes per day. These interventions provide extensive opportunities for children who have experienced some learning difficulties within the general classroom context to learn important skills. CRTIEC is continuing efforts to determine how to promptly identify children who are not making adequate progress during intervention and then design more intensive intervention that will support their learning.

Future research should focus on achieving stronger treatment effects. For the oral language domain, although effect sizes were large, children learned approximately half of the words taught. This effect is actually better than what is reported in other studies that explicitly taught challenging words to preschoolers (Jalongo and Sobolak 2011; Schwanenflugel et al. 2005). Factors that explain why the CRTIEC tier 2 interventions were more successful than other curricula include the large number of response opportunities provided in the intervention and the presentation of these opportunities in developmentally appropriate ways that actively engage young learners. Within the context of brief storybook reading sessions, the *My Story Friends* intervention has been shown to be successful. Because on average children typically learned only about half of the targeted words, greater intensity of instruction may be needed to produce stronger effects. Moreover, to produce robust improvements in later reading comprehension, it is likely that more opportunities for vocabulary learning will be necessary. Ideally, instruction would be expanded to incorporate vocabulary and comprehension lessons during other parts of the classroom day. The domain of comprehension remains a challenge, for both instruction and measurement. For example, relatively few opportunities can be embedded in stories to teach children to practice answering inferential comprehension questions compared to the number of opportunities that can be provided to teach children new vocabulary words.

Although research and development in tiered approaches in early literacy and language is in its infancy, some promising models have been developed. An early model by Gettinger and Stoiber (2007) is called the Exemplary Model of Early Reading Growth and Excellence (EMERGE). EMERGE involved the implementation of an entire three-tiered model focusing on the fundamental skills that have been empirically linked to reading success in the elementary grades: sound awareness, oral language, alphabet knowledge, and print awareness or SOAP. The lowest 50% of the class is identified for tier 2 intervention using the individual growth and development indicators. Children receive the intervention twice weekly in small teacher-led groups focused on SOAP-related skills integrated across thematic areas in the classroom curriculum.

Other more recently developed tier 2 curricular supplements (*Read It Again, Pre-K* (Justice et al. 2010), and *Developing Talkers* (Children's Learning Institute 2010)) provide teachers with specific guidance for individualized scaffolding as a means of differentiating instruction for children needing more support in early literacy and language. Preliminary information documenting the effectiveness of these interventions in improving early literacy skills such as rhyming, alliteration, print concepts, and alphabet knowledge as well as language skills have been reported (Justice et al. 2010; Zucker et al. in press). The approach used by CRTIEC for providing tier 2 intervention addresses these same skills but differs in its instructional design. More specifically, both the *My Story Friends* vocabulary/comprehension curriculum and the *PA Pathways to Literacy* curriculum incorporate design features to enhance their fidelity of implementation when implemented by teachers with less extensive early literacy or educational experience, such as paraprofessionals. Features such as the use of prerecorded stories with embedded learning trials or scripted lessons that provide explicit instruction and teacher feedback are ways that CRTIEC has attempted to increase opportunities to practice skills known to predict later success in reading.

Incorporating a multi-tiered approach to teaching in early childhood education settings is a recent phenomenon. The work undertaken thus far represents an initial step to ensure that teachers are offered tier 2 curricula that have been evaluated, refined, and shown to be effective. Future research must focus on how to promote proficiency and consistency in curricular implementation. Careful study must be carried out regarding the fidelity with which critical features of the intervention are implemented, how frequently the strategies are carried out, how well the strategies engage the young learner, and how each of these strategies relates to children's literacy and language outcomes. Moreover, if the goal is to produce strong effects in authentic education settings, practitioners, administrators, and policymakers will need to be members of the intervention development and implementation teams.

Tier 3 Intervention in Vocabulary and Early Literacy

Research Overview

Another major goal of CRTIEC was to develop and evaluate the effectiveness of interventions for preschool children who needed tier 3 supports in acquiring early literacy and language skills. Based on reviews of the research and intervention literature in these areas as well as an analysis of effective curricular and instructional design, the Center developed an early literacy intervention (*Reading Ready Early Literacy*) targeting phonemic awareness and alphabet knowledge. The goal of the intervention was to increase children's understanding of the alphabetic principle as well as their ability to associate sounds with letters and use sounds to read words. The focus of the tier 3 language intervention (*Reading Ready Language*) was the development of vocabulary and oral language skills. Both interventions consisted of brief (5–15 minute) teacher-led activities designed to be conducted one-on-one or in small groups in the preschool classroom. Instructional design of both interventions incorporated the following features:

1. Explicit teaching strategies
2. Priming of background knowledge
3. Ample opportunities to practice the targeted skills
4. Use of developmentally appropriate formats for teaching young children such as songs, finger plays, rhymes, and games
5. Use of print-related activities

The curriculum and its accompanying activities were piloted, evaluated, and revised through an iterative process involving a series of single-subject design studies focused specifically on the language or early literacy intervention.

Lessons Learned About Tier 3 Language and Early Literacy Intervention

To evaluate the effectiveness of the language intervention, a multiple baseline design across students was carried out in 11 prekindergarten classrooms with baseline consisting of "business as usual" vocabulary instruction delivered by the classroom teacher and intervention consisting of the *Reading Ready Language* instruction delivered by trained research staff. Children's progress was measured using a word knowledge measure assessing children's learning of targeted vocabulary words as well as growth in children's skill in describing targeted words in simple/expanded sentences. Over the 10-week intervention, 11 of the 18 students showed a change in level on the word knowledge measure with a 13.48 change (range = 1–37). Children also evidenced an average of six-point pre–post gains on standard scores on the *Clinical Evaluation of Language Fundamentals-5* (*CELF-5*; Wiig et al. 2004) and the *Peabody Picture Vocabulary Test* (*PPVT*; Dunn and Dunn 2007).

To evaluate the effects of the *Reading Ready Early Literacy* intervention, a multiple baseline across subjects study was conducted with ten children. Baseline was "business as usual" literacy instruction delivered by the classroom teacher and the intervention was the *Reading Ready Early Literacy* intervention implemented by research staff. To evaluate the effects of the intervention, *DIBELS Word Parts Fluency, First Sound Fluency* (*FSF*), and *Letter Naming Fluency* (LNF; Dynamic Measurement Group) were administered weekly. Results are summarized for FSF, the measure most closely aligned to the primary focus of the *Reading Ready* intervention. Visual inspection of the FSF data indicated a change in level for six of ten children and a change in trend for seven of ten. The average change in level for FSF was 9 (range = 2–18). Seven of the ten children achieved a score of ten correct sounds per minute by the end of the intervention, which is notable as ten correct sounds per minute is the benchmark goal at the beginning of kindergarten. Over the course of the 10-week intervention, children gained an average of 10.1 points on the Test of Preschool Early Literacy (TOPEL) phonemic awareness subtest (a 1 SD change; Lonigan et al. 2007) and an average of 3.5 points on the TOPEL print knowledge subtest.

Across studies, all students showed some gain in skills; however, effects were minimal for some students and large for others. These findings were not surprising given the diversity of skills and abilities in the children identified for tier 3. Some children were on individualized education plans (IEPs) for communication impairment, developmental delay, or behavior. Some children were dual language learners (DLL); others displayed severe challenging behavior. In general, children who were on IEPs made smaller gains compared to typically developing peers who received the intervention.

Children who were screened as eligible for tier 3 were a diverse group. Intervention for tier 3 support needs to be designed to provide enough structure for practitioners to implement with fidelity but at the same time, it must be flexible and responsive to individual needs of children. The amount of time dedicated to tier 3 intervention should be carefully considered and may be different for children who need early literacy compared to those needing tier 3 language support. A tier 3 intervention consisting of 5–15 minutes a day across 8–12 weeks may be enough time for increasing early literacy skills but may not be enough for language. Providing 5–15 minutes per day of one-to-one learning opportunities was important to provide opportunities for children with limited language to be exposed to and practice using new words. To be optimally effective, the tier 3 language intervention needs to be supplemented, with multiple opportunities for children to use language at their level throughout the day.

Some classrooms did not have a tier 1 focus on language and early literacy in place. This made identification of children in need of tier 3 intervention difficult. It is critical that before children are placed in intensive intervention, they have the opportunity to acquire skills through tier 1 instruction. Children may have shown delays in early literacy skills such as alphabet knowledge and identification of first sounds simply because they had limited exposure to these concepts (Burns et al. 1999; McGee and Richgels 1996). In addition, a number of classrooms did not have a classroom environment that supported small-group or one-to-one instruction. Tier 3 lessons took place within the classroom during "choice time" which could be distracting if children saw their peers engaging in what might be preferred activities. Finally, dosage of intervention may have influenced outcomes because of variability in children's attendance. Although the intervention was designed to occur three times a week for a period of 10 weeks for a total of 30 sessions, only one child received the full 30 sessions of intervention. The average number of intervention sessions was 20 with a range of 12–30. When examining gain scores of students who received more than 20 sessions versus those who received fewer than 20, the students who received more intervention made larger gains on all measures than students who received less intervention.

Implications of the Findings for Practice, Policy, and Research

There is not a standard RTI approach that will work in all early childhood settings. A model of RTI is needed that provides structure, a framework, and specific components (e.g., appropriate assessment tools, research-based interventions) that are adaptable for different populations and settings. Programs and classrooms need to be structured to support the kinds of activities that are necessary for successful RTI implementation (e.g., ongoing assessment, small-group activities for skill acquisition and practice). Professional development and training of early childhood staff will be essential in areas of language and early literacy skill development, assessment (including use of data), and instruction/intervention. To support a professional development model, additional resources and an ongoing system of support for early childhood professionals will be needed to implement an RTI model successfully.

Development of Measures for Identification and Progress Monitoring

Research Overview

A critical aspect of MTSS is a measurement framework that can be used to identify children who might benefit from additional levels of intervention and also to monitor growth in skill development for those children receiving increased levels of intervention. Through a process of iterative test construction and validation, CRTIEC developed and field-tested a measurement architecture to support instructional decision-making within an early childhood RTI framework using second-generation Individual Growth and Development Indicators (IGDIs).

IGDIs were originally developed to provide developmental assessment across all domains for children from birth to third grade. This work also led to a group of measures specifically targeting early literacy skills in preschoolers (McConnell and Missall 2008). The original IGDIs were part of a larger effort to develop program performance measures in early childhood special education (Priest et al. 2001). One shortcoming of this earlier version was that some of the measures proved to be too difficult for many low-performing students and frequently resulted in scores of zero. This shortcoming was problematic for RTI because discriminating levels of performance among low-performing students is a goal of RTI.

Under the auspices of CRTIEC, the original IGDIs for preschool children have been extensively redesigned and evaluated and are known as IGDIs 2.0. These measures have been developed to support the identification of students requiring supplemental or intensive intervention in the key early literacy domains of oral language, phonological awareness, alphabet knowledge, and comprehension. Research and development began by generating working definitions of the multiple subdomains of interest within early literacy based on recent reviews of language and early literacy development (e.g., National Early Literacy Panel 2009; Snow et al. 1998). Sets of items were developed after pilot efforts to test multiple formats for measuring child performance in these subdomains.

Item response theory (and, specifically, Rasch modeling) was used to construct sets of IGDI 2.0 items that closely aligned with the instructional decisions that needed to be made within RTI (Rodriguez 2010). As part of a larger model of RTI in early childhood programs, IGDIs 2.0 were developed to inform decisions about whether children attending preschool in the year before kindergarten demonstrated adequate levels of performance given the general level of instruction (i.e., tier 1), or if their performance indicated a need for more intense levels of instruction (i.e., tier 2 or tier 3).

Identification Identification IGDI scales were developed for each of three seasons of an academic year. Each Identification IGDI measure is

15 items long, with items selected to maximize the ability to detect children who were performing below a criterion cut point of desired development—and, thus, were candidates for higher-intensity intervention (Bradfield et al. in press; Wackerle-Hollman in press). These cut points were established through a standard-setting process using performance level descriptors (PLDs) and the contrasting group design (see Cizek and Bunch 2007). The contrasting group design was used to examine differences in teacher ratings of children's early literacy performance of children in their classes who were categorized as tier 1 (i.e., children making adequate progress toward end-of-preschool expectations given the typical level of instruction), tier 2 (i.e., children demonstrating less than expected levels of performance and were unlikely to achieve end-of-preschool expectations without some modest increase in intervention intensity), and tier 3 (those who were demonstrating less than expected levels of performance and were unlikely to achieve end-of-preschool expectations without some significant increase in intervention intensity). Children were then assessed on the IGDIs to determine the distributions of children in each of the performance levels (i.e., tiers). Using teacher-assigned PLD ratings to group children into tier categories and the actual IGDI distribution of children assigned to those categories, a cut point that best discriminated between tier 1 and tier 2 groups was identified (between Tier 1 and Tier 2; Rodriguez et al. 2011). IGDI 2.0 scores did not have adequate reliability to make fine-point distinctions between children who were identified by teachers as candidates for tier 2 or tier 3 intervention. Therefore, only a single cut score was identified for each measure: one that differentiated candidates for tier 1 versus tiers 2 and/or 3.

Validation studies with iterative test improvement and refinement were conducted during 3 years of CRTIEC. To obtain concurrent criterion validity evidence, each child participating in the validation studies completed one of two criterion measures: The CELF-5 (Wiig et al. 2004) or the TOPEL (Lonigan et al. 2007). Based on these studies, the technical adequacy of identification measures was established. Internal reliability estimates have ranged from 0.74 to 0.90 for the individual measures and concurrent construct validity coefficients have ranged from 0.49 to 0.71 (see Bradfield et al., in press; Wackerle-Hollman et al. in press). Classification accuracy is the real test of screening and identification measures; in other words, how accurately does a particular measure identify children that "truly" need an intervention, and those who "truly" do not? The challenge is significant because as a measure is adjusted to increase the identification of children needing additional intervention (*sensitivity* of the measure); it is generally the case that children who do not, in fact, need the intervention will also be identified (*specificity* of the measure). In many decision-making frameworks, including CRTIEC's, these "true-positive" and "false-positive" outcomes are balanced as best as possible, and additional procedures (progress monitoring, teacher nominations, and others) are used to adjust measure-only identification rates. In the CRTIEC work, classification accuracy was established with a focus on reducing false-negative decisions (i.e., failing to provide intervention to children who needed intervention) by holding sensitivity for cut scores to a minimum of 0.70 (see Jenkins et al. 2007). With this constraint, specificity levels varied with a mean of 0.56 across measures (Bradfield et al. 2013). In addition, examination of the correlations between the IGDIs and related field-recognized standardized measures of the early literacy and language constructs provided strong evidence of the IGDIs construct-related validity.

The one IGDI measure that failed to show adequate construct validity was Rhyming 2.0. The correlation between Rhyming 2.0 and the *TOPEL* phonological awareness subtest ($r=0.49$), revealed only a moderate relation between scores on these two measures. The hypothesis for this finding was that the TOPEL PA included very few rhyming items, and tapped the construct of phonological awareness in a much broader manner than the more constrained skill of rhyming. These results may represent one of the challenges in language and literacy assessment and intervention within an RTI model; as the focus of assessment and intervention shifts from constrained to unconstrained skills (c.f., Paris 2005), relations to both concurrent validity measures and sensitivity to intervention appear affected. These issues require additional analysis and research to fully describe the utility of measures.

Progress Monitoring One of the central tasks of any RTI model is to identify individual children who, based on current level of achievement, might benefit from more intensive, intervention (typically, tier 2 or tier 3 in RTI). After children begin receiving a more intensive tier of instruction, it is important to monitor how their skills are growing in that domain at frequent and consistent time points. These data are central to determining if the added intensity of intervention is contributing to individual child growth, or if higher (or lower) levels of intervention intensity would be appropriate (see Barnett et al. 2007).

To meet the need for progress monitoring in early childhood RTI models, specialized scales of all five IGDIs were developed to be administered every 3 weeks to children receiving tier 2 or 3 early literacy and language interventions. These progress-monitoring IGDIs are made up of 30 items; all 30 items measure ability levels at the *lower range* of child performance, so that these progress monitoring scales are maximally sensitive to changes in performance for tier 2 and tier 3 students (i.e., lower-performing at-risk children). Field trials are currently being conducted to evaluate the effectiveness of the progress-monitoring IGDIs for measuring growth over brief periods of time (3 and 6 weeks) and to examine their psychometric properties when used in this way. Preliminary data suggest that the progress-monitoring IGDIs 2.0 do appear to be sensitive to growth for children receiving tier 2 intervention.

Implications for Practice, Policy, and/or Research

For any systems-level framework to work effectively, the components of the framework need to be strong and rigorous and effectively support one another. This assessment package has been uniquely designed to support the language and literacy instructional decisions required within an early childhood RTI framework. Moving forward, it will be important to examine the degree to which the information provided by the assessments aligns with and supports use of specific language and literacy interventions. Specifically, programs must determine how to use the information provided by the identification IGDIs to plan tier 2 and 3 intervention.

Data-Based Decision-Making

Research Overview

While early educators have long been encouraged to collect and use student performance data to plan and evaluate instruction (Bagnato et al. 2010; McLean et al. 2003), RTI presents a new level of formal and rigorous requirements for collecting and using student-level data in prekindergarten programs (Greenwood et al. 2011; Greenwood et al. 2008). Within an RTI framework, data-based decision-making, or DBDM, is conceptualized as a set of procedures that are more systematic than those typically used in pre-K programs for making instructional decisions. Within RTI models, DBDM procedures are established for identifying children who may benefit from different levels of intervention, and in turn, evaluating children's responses to that intervention, to determine whether to continue or alter the current level of intervention services. DBDM in early childhood RTI models supports two different levels of decision making: identification and progress monitoring. First, following universal screening, DBDM assists teachers in identifying which children in a particular classroom or group might benefit from more intensive intervention. DBDM also provides a set of steps for collecting and interpreting progress-monitoring data for children in tier 2 or tier 3 services; progress data and associated DBDM algorithms help determine "hold or move" decisions for individual students, identifying instances where the intensity of tiered intervention should be changed to a different tier to most efficiently achieve medium- and long-term goals.

Figure 1 presents a simplified version of the CRTIEC DBDM model. In this model, two phases of assessment—universal screening for

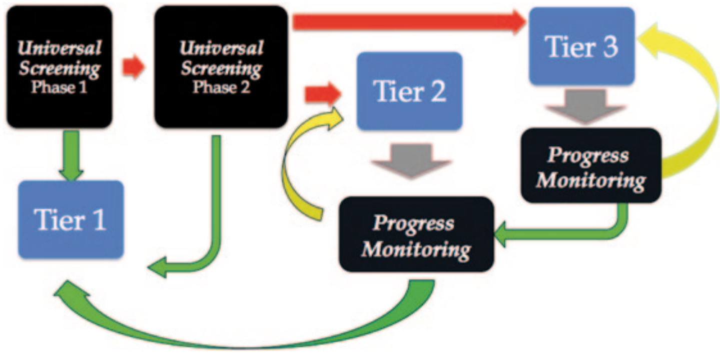


Fig. 1 CRTIEC data-based decision-making model. CRTIEC Center on Response to Intervention in Early Childhood

identification (typically completed in Fall, Winter, and Spring of each academic year) and progress monitoring (typically completed every 2–4 weeks during tiered intervention services)—are used to support all decisions. Green arrows indicate points of assessment where individuals meet expected levels of performance, and red arrows indicate decisions when individuals do not meet expected levels of performance; yellow arrows in progress monitoring indicate a child is making adequate progress and remain candidates for that level of intervention

Universal screening occurs in two phases, first administering IGDIs to all children, and then following up with teacher questionnaires or additional information; children who meet or exceed desired performance levels in either phase are considered tier 1 candidates, and children not meeting these performance standards are considered candidates for tier 2 or tier 3 intervention, depending on performance in both phases of screening. Feasibility of each of these components has been tested in community-based preschools. Reports from participating teachers indicate that individual child assessment is feasible and produces actionable data that teachers find useful (Potter et al. 2012).

Lessons Learned About DBDM

Conducting research to produce the components of this DBDM and to implement significant portions in research and applied settings has produced several important lessons. First, it is challenging to create sharp distinctions among either the tiers of intervention or the children who are best candidates for these various tiers. One of the challenges has to do with the psychometric properties of the measures themselves. Universal screening measures often have moderate standard errors of measurement, making determination of "above or below a benchmark" hard to complete with precision. But beyond the psychometric considerations, another challenge is the degree of ambiguity in research and practice about the particular child characteristics that best indicate need for tier 2 or tier 3 services. Specifically, while relative skill level is both necessary and sufficient to differentiate these children, other child characteristics, such as independent work behaviors, may influence the match between tiered interventions and individual children. In addition, RTI implies a level of differentiated instruction and intervention that is not yet commonplace in many early childhood classrooms, making tier determination less functional because of the lack of appropriate interventions after identification. Once early education programs have a set of valid tier 2 and tier 3 interventions, identifying good candidates for children who would benefit from those interventions will be less ambiguous than currently. The availability of valid tier 2 interventions also will allow programs to identify children for tier 3 who fail to demonstrate progress in response to tier 2 strategies. Finally, a significant challenge for RTI in pre-K programs is that they have a limited number of months to find the children who need additional support and implement effective intervention to get those children ready to learn to read in kindergarten. As such, pre-K RTI programs have an urgency in identifying children who would benefit from additional tiers of intervention, implement them with adequate dosage and fidelity, and determine if they are changing children's trajectories of learning the skills they will need to succeed in kindergarten.

RTI and DLL

While defining how RTI might be applied in early childhood settings is a significant task all on its own, an added dimension, defining RTI for DLL, makes for increased challenges and opportunities. Across the nation, early childhood professionals and classrooms are encountering increasing numbers of DLL students. The largest and most prevalent subgroup within DLLs is Spanish-speaking children, including families and children from Latino and Hispanic backgrounds. Over the past decade, the percentage of Latino children who attend US schools has increased dramatically, approaching 25 % in the year 2007 (Garcia and Jensen 2009). Further, Latinos specifically are overrepresented in the category of learning disabilities (LD) with 56 % of all Latino students in special education programs identified as LD with reading problems as their primary need for support (Waitoller et al. 2010; Zehler et al. 2003) and in the nonproficient performance categories of the National Assessment of Educational Progress (Foorman, this volume; Soci & Vanderwood, this volume). Spanish-speaking DLLs are a particularly important research target not only because they face the typical challenges and opportunities present in early childhood classrooms but also because their classroom success is influenced by their *level and rate of performance* across two languages (native and secondary). Their classroom performance is further influenced by their teachers' understanding of how learning two languages may contribute or hinder academic success (Paradis et al. 2010; Peña et al. in press, Zepeda et al. 2011). To contribute to ongoing research efforts to empirically investigate the best method for employing RTI with DLL students, an analysis of data from DLL subgroups has been carried out from two ongoing CRTIEC studies to examine DLL students' level and rate of performance. First, from the large study of tier 1 described earlier, an analysis was conducted of a subsample of DLL students' performances on the IGDI 2.0 screening/identification measures to examine trends in performance that may be unique to DLL students. Second, an examination of item level responses on each item within IGDI 2.0 measures was carried out using differential item functioning (DIF). DIF allows for exploration of response trends that may be unique to an identified subgroup (Thissen et al. 1993). Taken together these findings contribute to the growing research base on best practices for DLL students within an RTI model.

Lessons Learned about DLL in Tier 1

To understand more about the diverse group of children identified as DLL and how they were performing in early literacy and language when they received only tier 1 support in these areas, an examination was carried out during the large tier 1 study described earlier of level and rate of performance of children identified as DLLs compared to English-only (EO) students. In the larger study, a sample of 601 4-and 5-year-old children was recruited: 505 EO preschoolers (81 %) and 96 DLL preschoolers (19 %). Students participated during the year prior to kindergarten across four regions of the USA (Missouri/Kansas, Ohio, Oregon, and Minnesota). Within this sample, students were gathered from 65 classrooms (39 % from state-funded prekindergarten programs, 31 % from Title 1, 23 % from Head Start, and 7 % from private tuition preschools).

To examine performance in the language of instruction in these programs (English), trained administrators gave all students the *sound identification* and the *picture naming* IGDI 2.0 measures three times annually (fall, winter, and spring). Results indicated that DLL students' levels and rates were significantly different from those of EO students for both picture naming and sound identification. On picture naming (the measure of oral language performance), DLL students scored significantly lower than their EO counterparts but grew faster. DLLs scored 14.2 points lower than EO children at the beginning of the year (fall) and 9.8 points lower than EO children in the spring (September t (503) = −13.03, $p < 0.001$; May t (503) = −9.73, $p < 0.001$). Results for *picture naming* also indicated that DLL students grew at a significantly faster rate than EO children (t (1406) = 6.73, $p < 0.001$).

For sound identification, DLL students also started the year with significantly lower levels of performance than their EO counterparts (September t (503) = −3.38 $p < 0.005$; May t (503) = −0.93 $p = 0.352$). However, DLL students were able to close this gap by the end of the year, with no significant difference between DLL and EO students on *sound identification* in spring. In addition, rate of performance for DLL students was significantly different from EO peers such that the DLL students grew faster, allowing them to reduce the gap in performance by year's end (t (1404) = 2.12, $p < 0.05$).

In a second analysis relevant to DLL children, item-level responses of DLL children were compared to the other children to determine the cultural appropriateness of each item on the IGDI 2.0 measures. Specifically, an examination was carried out to describe how each item within the IGDI 2.0 measures (*picture naming, rhyming, first sounds, which one doesn't belong,* and *sound identification*) contributed to the task and provided unique information (if any) for DLLs. Using the item response theory statistics, DIF analyses were conducted on items across each measure. Items that were administered to more than 30 students who were considered DLLs were evaluated. A few items yielded significant DIF on most of the IGDI 2.0 measures. For example, in the *sound identification* measure, items that were responded to differently by the DLL children included items that contained letters not frequently observed in the Spanish language (i.e., K and W). For *picture naming,* 14 items demonstrated DIF. Of these 14, seven items demonstrated DIF that favored DLL student performance (nails, comb, doll, deer/fawn, gate/fence, broom, and watch). The remaining seven items did not favor DLL student performance (pears, flute, map, egg, shell, chef/cook, lock, and violin). In some cases, construct-irrelevant hypotheses were identified for response patterns. For example, for the image of the chef/cook the image is of a man. DLL students may answer incorrectly on this item because they may have a strong cultural influence that suggests women are cooks. These items were excluded from the IGDI 2.0 screening/identification sets. The end result of the refinement of IGDI 2.0 measures based on this item analysis is a more culturally sensitive measure that can evaluate English acquisition in children who have Spanish as their first language

Implications of the Findings for Practice, Policy, and Research

Research findings from investigations of dual language development and best practice recommendations in assessment suggest that measuring bilingual children in both languages yields the most accurate results for decision-making within an RTI model (Peña and Halle 2011). Recent research has indicated that DLLs have two separate language systems (Paradis et al. 2010). Specifically, while the single-language vocabularies of young DLLs are usually smaller than monolingual children, their total vocabulary across both languages is similar to monolingual peers (Hammer et al. in press). It is critical then, that measures developed in languages other than English pay careful attention to differences in linguistic structure so that information from each language can be used to best inform instruction and intervention practices.

At the same time, there is a paucity of sound methods to identify which students are DLLs. Moving forward in the field to serve all students, careful attention to how we define what constitutes DLL is warranted. In addition, there is a considerable amount of diversity within the DLL population. When assessing children's needs (i.e., trying to decide if they could benefit from a tier 2 or tier 3 intervention), it is important to distinguish between problems associated with low English language skills and other issues.

States' Progress in Implementing RTI in Early Childhood

One final aspect of the work scope of CRTIEC has been to assess how extensively RTI is being implemented nationally in preschool and Head Start classrooms. To fulfill this mission, CRTIEC has conducted an annual survey of state-level personnel since 2009. State-level early childhood leaders including IDEA-Part B [619] directors, state pre-K directors, and Head Start State Collaboration Office Directors have been surveyed for the purpose of examining the reported implementation status of RTI in early childhood settings across US states and territories.

In addition to assessing the extent to which RTI is implemented nationally, trends in degree of implementation over time across the states and identifying challenges perceived by state leaders to inhibit or delay implementation of MTSS are also examined. Among the goals of this study is the identification of RTI components being used across the states, the curricula and learning outcomes being emphasized, and the types of early childhood programs in which RTI is most frequently employed. Findings from this body of information indicate a number of national trends in RTI practices across early childhood settings.

The number of states and territories responding to the survey has increased each year, which may imply growing interest in RTI as a method of both identifying young children whose needs may exceed what is currently offered them, and as a method of providing additional support to meet their needs. Over the past 4 years, there has been a 16 % increase in the number of states reporting local implementation of RTI in preschools. In 2013, 47 % of states reported that RTI approaches were being implemented at least in some centers within their locales. Further survey results indicate that professional development related to RTI continues to take place. This level of commitment is very promising as few states have as yet adopted state-wide policy related to RTI in early childhood.

Survey respondents have indicated that RTI in preschools is primarily targeting early literacy and social/behavioral outcomes. Of the 45 US states and territories responding to this question in the most recently completed survey, 73 % identified early literacy, 69 % pinpointed social emotional, 42 % of states included numeracy, and 36 % identified language. Finally, survey respondents have pointed to a number of significant challenges and concerns in implementing MTSS. Most frequently reported significant challenges included insufficient opportunities for training staff to implement RTI components, and limited resources necessary to develop the infrastructure to support an RTI model. Further noted were the lack of evidence-based tier 1 programs, tier 2 and tier 3 interventions, and progress-monitoring measures.

Ongoing results from this study suggest several potentially important issues that may lead to future research. First, greater collaboration across programs and stakeholders within and across states must occur to align the terminology used in MTSS to allow for clearer and unambiguous conversations about RTI. Respondents' comments throughout this research frequently serve to clarify the language used to describe RTI in their states and the conflicting policies that present roadblocks to collaborations. The second item worthy of consideration is a need for policy, and policy clarification. That individuals in state-level leadership positions find survey completion challenging as a result of unavailable or unclear information about RTI speaks to a prescient need for clear and direct policy related to the planning and implementation of early childhood RTI in states where none is currently present. Further, there is a need for the consistent designation of a state-level individual responsible for state-wide dissemination and implementation of RTI-related policy and activities. Finally, states, programs, and school districts must all determine their future direction in terms of implementation of an RTI model in early childhood education settings and specifically identify how, where, and for what ultimate purpose RTI in early childhood will be employed.

In summary, RTI and other MTSS are increasingly employed in early childhood settings across US states and territories. Most states report implementing evidence-based tier 1 and having developed collaborative relationships to facilitate this implementation. The vast majority of states report not yet having RTI models that can be shared with others. However, some states report that some local districts have been implementing RTI successfully for many years. These data suggest that some select early childhood programs have information that can be shared about RTI and MTSS and the features that are working. While the early childhood field is still at the beginning stages of implementing RTI, the potential for broader dissemination is there as multiple states and districts are looking for ways to meet the diverse needs of young children in their programs.

Conclusions

The work described in this chapter represents the authors' attempt to address the issue of large numbers of children entering schools already at risk for learning to read. The work shows that significant percentages of children can be identified in their prekindergarten year as showing significant delays in early literacy and language. What is disturbing however is that these children are typically not receiving much instruction focused on early literacy and language and are therefore on a trajectory of being unprepared for kindergarten expectations.

The work carried out in this Center provides a framework for changing the way we approach children at risk for early literacy problems. A method for identifying those children who would benefit from additional instructional support in early literacy at the beginning of their prekindergarten year is described. A set of interventions that could be provided for those children needing some additional support in the form of targeted interventions (tier 2) and for those needing considerably more support in the form of tier 3 interventions is provided. What is described in this chapter is the work carried out to show the efficacy of these interventions in improving children's early literacy and language performance. Although some work has demonstrated that these interventions can be carried out by teachers in authentic preschool settings, more work is needed in this area.

A number of challenges still remain before Multi-tiered Systems of Support are ready for prime time in early education settings. One of the biggest issues concerns the need for a greater focus on higher quality and greater quantity of instruction in early literacy in preschool programs. Too many programs had no curriculum with a specific emphasis on literacy and language and relatively little professional development in these areas. Therefore, little change in children's performance relative to typically developing peers as a result of prekindergarten instruction was seen. If programs are serious in terms of getting children ready for school, they need to raise the quality of instruction for all children. Then, those children who need more than tier 1 evidence-based literacy instruction can be identified and receive enhanced instruction in tier 2 or tier 3 interventions.

The work described in this chapter shows that tier 2 and tier 3 interventions can improve children's school readiness trajectories but they will be challenging to implement in classrooms where more than 30% of children need this additional support. Clearly, more work is needed on curriculum development and the system reform and professional development required to sustain high-quality instruction for children at all performance levels. RTI offers opportunities in early childhood to improve assessment, intervention, and professional development practices, and ultimately the learning trajectories of all children.

References

Atwater, J. B., Lee, Y., Montagna, D., Reynolds, L. H., & Tapia, Y. (2009). *Classroom CIRCLE: Classroom code for interactive recording of children's learning environments (Version 2.0)*. Kansas City: Juniper Gardens Children's Project, University of Kansas.

Bagnato, S. J., Neisworth, J. T., & Pretti-Frontczak, K. (2010). *LINKing authentic assessment and early childhood intervention: Best measures for best practices* (2nd ed.). Baltimore: Brookes.

Barnett, D. W., VanDerHeyden, A. M., & Witt, J. C. (2007). Achieving science-based practice through Response to Intervention: What it might look like in preschools. *Journal of Educational and Psychological Consultation, 17*(1), 31–54.

Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2012). *The state of preschool 2012: State preschool yearbook*. New Brunswick: National Institute for Early Education Research.

Bradfield, T. A. (2013). [Results of the Decision Making Framework Analysis]. Unpublished raw data.

Bradfield, T. A., McConnell, S. R., Rodriguez, M., & Wackerle-Hollman, A. (2013). A decision-making framework for early language and literacy early childhood response to intervention. Unpublished technical report, Center for Response to Intervention in Early Childhood, University of Minnesota.

Bradfield, T., Besner, A., Wackerle-Hollman, A., Albano, A., Rodriquez, M., & McConnell, S. (2014). Redefining individual growth and development indicators: Oral language. *Assessment for Effective Intervention, 73,* 233–244.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6,* 2–11.

Burns, M. S., Griffin, P., & Snow, C. E. (Eds.). (1999). *Starting out right: A guide to promoting children's reading success*. Washington, DC: National Academy Press. http://www.nap.edu/html/sor/.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analysis of response-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Buysse, V., & Peisner-Feinberg, E. (2010). Recognition & response: Response to Intervention for PreK. *Young Exceptional Children, 13*(4), 2–13.

Buysse, V., & Peisner-Feinberg, E. S. (2013). Response to intervention: Conceptual foundations for the early childhood field. In V. Buysse & E. S. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 3–23). Baltimore: Brookes.

Carta, J. J., & Greenwood, C. R. (2013). Promising future research directions in response to intervention in early childhood. In V. Buysse & E. S. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 421–432). Baltimore: Brookes.

Chandler, L. K., Miller Young, R., Nylander, D., Shield, L., Ash, J., Bauman, B., Butt, J., et al. (2008). Promoting early literacy skills within daily activities and routines in preschool classrooms. *Young Exceptional Children, 11*(2), 2–16.

Children's Learning Institute. (2010). *Developing talkers: Pre-K curricular supplement to promote oral language*. Houston: University of Texas Heath Science Center.

Cizek, G. J., & Bunch, M. B. (Eds.). (2007). *Standards for educational psychological testing*. Washington, DC: AERA.

Dickinson, D. K., & Tabors, P. O. (Eds.). (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore: Brookes.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test—fourth edition*. Bloomington: Pearson Assessments.

Ezell, H. K., & Justice, L. M. (2005). *Shared storybook reading: Building young children's language and emergent literacy skills*. Baltimore: Paul H. Brookes.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30–37.

Garcia, E., & Jensen, B. (2009). *Early educational opportunities for children of Hispanic origins* (Social Policy Rep. No. 23-2). Ann Arbor: Society for Research in Child Development.

Gettinger, M., & Stoiber, K. (2007). Applying to response-to-intervention model for early literacy developing in low-income children. *Topics in Early Childhood Special Education, 27,* 198–213.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2013). DIBELS first sound fluency. In DIBELS next technical manual. Eugene: Dynamic Measurement Group Inc.

Greenwood, C. R., Kratochwill, T. R., & Clements, M. (Eds.). (2008). *Schoolwide prevention models: Lessons learned in elementary schools*. Baltimore: Guilford Press.

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011a). The Response to Intervention (RTI) approach in early childhood. *Focus on Exceptional Children, 43*(9), 1–22.

Greenwood, C. R., Carta, J. J., Atwater, J., Goldstein, H., Kaminski, R., & McConnell, S. R. (2012). Is a Response to Intervention (RTI) approach to preschool language and early literacy instruction needed? *Topics in Early Childhood Special Education, 33,* 48–64.

Gutiérrez, K. D., Zepeda, M., & Castro, D. C. (2010). Advancing early literacy learning for all children: Implications of the NELP report for dual-language learners. *Educational Researcher, 39*(4), 334–339.

Hammer, C. S., Gillanders, C., Hoff, E., Uchikoshi, Y., & Castro, D. C. (2014). *Language and literacy development in dual language learners: A critical review of the research Early Childhood Research Quarterly, 29,* 715–733.

Jalongo, M. R., & Sobolak, M. J. (2011). Supporting young children's vocabulary growth: The challenges, the benefits, and evidence-based strategies. *Early Childhood Education Journal, 38,* 421–429.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36,* 582–600.

Justice, L. M., & Kaderavek, J. N. (2004). Embedded-explicit emergent literacy intervention I: Background and description of approach. *Language, Speech, and Hearing Services in Schools, 35,* 201–211.

Justice, L. M., McGinty, A., Cabell, S. Q., Kilday, C. R., Kighton, K., & Huffman, G. (2010). Language and literacy curriculum supplement for preschoolers who are academically at risk: A feasibility study. *Language, Speech, and Hearing Services in Schools, 41,* 161–178.

Kelley, E. S., Goldstein, H., Spencer, T. D., & Sherman, A. (2014). An automated tier 2 vocabulary and comprehension intervention for preschool children with limited oral language skills. Unpublished manuscript under review.

Lonigan, C., & Wilson, S. B. (2008). Report on the *Revised Get Ready to Read! Screening Tool*: Psychometrics and normative Information. Final Technical Report Prepared for the National Center on Learning Disabilities.

Lonigan, C., Wagner, R., Torgesen, J., & Rashotte, C. (2007). *The Test of Preschool Early Literacy* (*TOPEL*). Austin, Pro-Ed.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335.

McConnell, S. R., & Missall, K. N. (2008). Best practices in monitoring progress for preschool children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 561–573). Washington, DC: National Association of School Psychologists.

McCollum, J. A., Hemmeter, M. L., & Hsieh, W.-Y. (2013). Coaching teachers for emergent literacy instruction using performance based feedback. *Topics in Early Childhood Special Education, 33,* 28–37.

McGee, L. M., & Richgels, D. J. (1996). *Literacy's beginnings: Supporting young readers and writers* (2nd ed.). Boston: Allyn and Bacon.

McGinty, A. S., Justice, L. M., Piasta, S. B., Kaderavek, J., & Fan, X. (2012). Does context matter?: Explicit print instruction during reading varies in its influence by child and classroom factors. *Early Childhood Research Quarterly, 27,* 77–89.

McLean, M., Wolery, M., & Bailey, D. B. (2003). *Assessing infants and preschoolers with special needs* (3rd ed.). New York: Prentice Hall.

National Early Literacy Panel. (2009). *Developing early literacy: Report of the National Early Literacy Panel—A scientific synthesis of early literacy development and implications for intervention*. Jessup: National Institute for Literacy.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4,* 116–135.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). Child-care structure—process—outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13,* 199–206.

Noe, S., Spencer, T. D., Kruse, L., & Goldstein, H. (2013). Effects of a tier 3 phonological awareness intervention on preschoolers' emergent literacy. *Topics in Early Childhood Special Education* 0271121413489172, first published on June 6, 2013 as doi:10.1177/0271121413489172.

Paradis, J., Genesee, F., & Crago, M. (2010). *Dual language development and disorders: A handbook for parents and professionals*. Baltimore: Brookes.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78*(4), 1255–1264.

Peña, E. D., & Halle, T. (2011). Assessing preschool English learners: Traveling a multi-forked road. *Child Development Perspectives, 5*(1), 28–32.

Peña, E. D., Gillam, R. B., Bedore, L. M., & Bohman, T. (2011). Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. *American Journal of Speech Language Pathology, 20,* 302–314.

Pew Charitable Trusts. (2004). *Percent of population ages 3 and 4 who are enrolled in school: Census 2000*. http://www.pewtrusts.org/news_room_detail.aspx?id=19570. Accessed 4 July 2013.

Pianta, R. C., LaParo, K. M., & Hamre, B. (2008). *Classroom assessment scoring system*. Baltimore: Brookes Publishing.

Pierce, C. D., & Bruns, D. A. (2013). Aligning components of Recognition and Response and Response to Intervention to improve transition to primary school. *Early Childhood Education Journal, 41,* 1–8.

Potter, L., Besner, A., Palma, J., Bradfield, T., & McConnell, S. (2012, September). Using a partial-credit model with Picture Naming IGDIs 2.0. Poster presented at the Fourth Annual Summit on Response to Intervention in Early Childhood, Santa Ana Pueblo NM.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness* (*NCER 2008-2009*). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education, U.S. Government Printing Office.

Priest, J. S., McConnell, S. R., Walker, D., Carta, J. J., Kaminski, R. A., McEvoy, M. A., Good, R. H., Greenwood, C. R., & Shinn, M. (2001). General growth outcomes for young children: Developing a foundation for continuous progress measurement. *Journal of Early Intervention, 24,* 163–180.

Rodriguez, M. C. (2010). *Building a validity framework for second-generation IGDIs*. University of Minnesota. Minneapolis.

Rodriguez, M. C., Albano, A. D., McConnell, S., Wackerle-Hollman, A., & Bradfield, T. (April, 2011). Standard setting with innovative measures of early literacy: Contrasting groups. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Schwanenflugel, P. J., Hamilton, C. E., Bradley, B. A., Ruston, H. P., Neuharth-Pritchett, S., & Restrepo, M. A. (2005). Classroom practices for vocabulary enhancement in prekindergarten: Lessons from PAVEd for success. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 155–178). Mahwah: Lawrence Erlbaum.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Spencer, E., Goldstein, H., Sherman, A., Noe, S., Tabbah, R., Ziolkowski, R., & Schneider, N. (2012). Effects of an automated vocabulary and comprehension intervention: An early efficacy study. *Journal of Early Intervention, 34,* 195–221.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. Holland (Ed.), *Differential item functioning* (pp. 67–113). Hillsdale: Erlbaum.

Vandell, D. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly, 50,* 387–414.

Wackerle-Hollman, A. (2013). [Results of Progress Monitoring Analysis]. Unpublished raw data.

Wackerle-Hollman, A., Schmitt, B., Bradfield, T., Rodriguez, M., & McConnell, S. (In press. Online November 14, 2013). Redefining individual growth and development indicators: Phonological awareness. *Journal of Learning Disabilities*. doi:10.1177/0022219413510181.

Waitoller, F. R., Artiles, A. J., & Cheney, D. A. (2010). The miner's canary: A review of overrepresentation research and explanations. *The Journal of Special Education, 44,* 29–49.

Wiig, E., Secord, W., & Semel, E. (2004). *Clinical evaluation of language fundamentals-preschool* (2nd ed.). San Antonio: Harcourt Assessment.

Zehler, A. M., et al. (2003). *Descriptive study of services to LEP students and LEP students with disabilities*. Final Report to the U.S. Department of Education, Office of English Language Acquisition. Arlington, VA: Development Associates, Inc.

Zepeda, M., Castro, D. C., & Cronin, S. (2011). Preparing early childhood teachers to work with young dual language learners. *Child Development Perspectives, 5*(1), 10–14.

Ziolkowski, R. A., & Goldstein, H. (2008). Effects of an embedded phonological awareness intervention during repeated book reading on preschool children with language delays. *Journal of Early Intervention, 31,* 67–90.

Zucker, T. A., Solari, E. J., Landry, S. H., & Swank, P. R. (2013). Effects of a brief, tiered lanaguage intervention for pre-kindergarteners at risk. *Early Education and Development,* 366–392.

Teacher Formative Assessment: The Missing Link in Response to Intervention

Linda A. Reddy, Christopher M. Dudek and Elisa S. Shernoff

Response to intervention (RTI) is an education framework emphasizing multi-tiered systems for prevention, identification, and intervention for *all* students at risk for learning and behavioral challenges. RTI organizes educational resources and service delivery for general and special education into an integrated problem-solving approach for promoting student achievement (Batsche et al. 2005; [NCRI] National Center on Response to Intervention 2010). Successful implementation of RTI requires collaboration between parents, educators, related service providers, and administrators to promote students' success (Jackson et al. 2009). RTI also provides school districts with an organizational framework for promoting student achievement, providing immediate learning supports as needed, reducing the number of special education referrals, and enhancing classroom instruction (Jackson et al. 2009).

Within an RTI framework, educators use progress-monitoring tools to adapt instruction to meet diverse student learning needs. While RTI has several models, consensus exists among scholars on the core principles and overall framework (e.g., Fuchs and Fuchs 2006; Klinger and Edwards 2006; Mellard et al. 2009; Jackson et al. 2009). National organizations, coalitions, and researchers have defined the key features of an RTI model to include: (a) utilizing a research-based core curriculum, (b) universal screening, (c) a multi-tiered system of evidence-based interventions matched to student need, (d) data-based decision-making regarding instructional practices, and (e) progress monitoring through formative and summative assessments (e.g., Batsche et al. 2008; Chun and Witt 2008; Fuchs and Fuchs 2006; Mellard et al. 2009). These principles are woven into a three-tiered service and resource delivery system (Fuchs and Fuchs 2006; Klinger and Edwards 2006).

Although RTI models have traditionally focused on identifying and supporting students at risk for learning and behavioral difficulties, recent models emphasize supporting *all* learners in school systems (Bender and Shores 2007; National Association of State Directors of Special Education 2007; Jackson et al. 2009). The recent emphasis on supporting all learners is critical given RTI's potential impact on the important proximal and distal factors affecting student learning; in particular, the quality of education students receive. A key feature across RTI models includes the provision of evidence-based, high-quality

Author Note: The research reported here was supported by the Institute of Education Sciences, US Department of Education, through Grant R305A080337, Brady Education Foundation, and US Department of Education–Teacher Incentive Fund Program, through Grant S374A120060 to Rutgers University. The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education. Correspondence concerning this chapter should be addressed to Linda A. Reddy, Ph.D. Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ. 08854; E-mail: LReddy@scarletmail.rutgers.edu.

L. A. Reddy (✉) · C. M. Dudek · E. S. Shernoff
Rutgers University, New Brunswick, USA
e-mail: LReddy@scarletmail.rutgers.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_34

instructional practices linked to positive student learning outcomes (e.g., Fuchs and Fuchs 2006; Klinger and Edwards 2006; Griffiths et al. 2007). Likewise, RTI has been defined as providing high-quality instruction, matched to student need, that integrates ongoing assessment of performance to make educational decisions (Batsche et al. 2005).

The emphasis within RTI on evidence-based instruction is warranted given the corpus of literature documenting the link between effective instructional practices, curricula, and student learning (National Reading Panel 2000; National Mathematics Advisory Panel 2008). Beyond curricula, there exist more than 60 years of literature documenting effective instructional practices and the general features of effective instruction that promote student learning (e.g., Bennet 1988; Creemers 1994; Good and Brophy 1980; Harris 1998; Hattie et al. 1996; Marzano 1998; Scheerens 1992; Walberg 1986; Wang 1991; Wang et al. 1993). Thus, effective instruction is a core tenet in the three tiers of an RTI model (i.e., tier 1/universal interventions, tier 2/secondary interventions, tier 3/tertiary interventions) to maximize academic achievement for all learners.

Combining RTI's focus on effective instruction with its foundational principles calls for highly qualified educators who possess the knowledge and skills to teach students in all tiers of the system (Barnett 2004; Espinosa 2002, National Research Council 2001). Teacher effectiveness has been documented as one of the most important variables influencing student learning and numerous studies have demonstrated that differential levels of effectiveness impact student achievement (e.g., Connor et al. 2005; Darling-Hammond 2000; Leigh 2010; Nye et al. 2004; Rivkin et al. 2005; Rockoff 2004; Sanders 2000). The necessity of effective teachers for implementing RTI models is further underscored by the reauthorization of No Child Left Behind and IDEA, which called for highly qualified teachers in every classroom (IDEA 2005). National attention on promoting effective teachers continues to grow in response to increasing RTI program implementation, as well as federal and statewide teacher evaluation and accountability initiatives (Bender and Shores 2007; Jimerson et al. 2007).

Despite the role that high-quality instruction plays in the implementation of RTI, prevailing models continue to emphasize student assessment, student progress monitoring, and student outcomes to the exclusion of support for teachers. Limited attention has been devoted to applying RTI principles to improving teacher competencies and enhancing teacher effectiveness (e.g., Crawford et al. in press; Pianta et al. 2011; Reddy et al. 2013; Reddy and Dudek 2014). The organizational framework of RTI assists school districts in promoting and managing student achievement, yet systems for enhancing teacher effectiveness have been historically absent in school districts.

The popular *Rush to Judgment* report criticized current teacher evaluation systems for being "superficial, capricious" and often failing to address instructional quality, professional development, and student outcomes (Toch and Rothman 2008, p. 1). The *Widget Effect* similarly concluded that modern educator evaluator systems failed to accurately capture differences in teachers' instructional performance or identify teachers' professional development needs (Weisberg et al. 2009). Researchers and practitioners have similarly identified common themes of failure that include limited practical feedback, limited utility in improving instructional practices, and a lack of shared understanding regarding effective teaching process (e.g., Danielson and McGreal 2000; Gersten et al. 2000; Heneman and Milanowski 2004; Johnson 1990; Marzano et al. 2011; Medley and Coker 1987). Current teacher evaluation systems also do not support or collect (valid) information on specific teacher practices (Reddy et al. 2015) and instead offer qualitative information typically used in summative assessments.

With the success of RTI systems hinging on high-quality instruction, current educator evaluation processes and systems must emphasize ongoing, specific feedback that supports planning and implementation of effective instructional and classroom behavioral management practices. Educators need information about specific evidence-based effective instructional practices, how best to implement these practices in their classroom, and how to monitor their effectiveness (Reddy et al. 2013a; 2013b, 2013c). What educators need is an *RTI that enhances their professional practices*.

Applying RTI principles to teacher evaluation offers a solution to current system failures. Promoting effective teaching, akin to student learning, requires the integration of assessment and intervention through ongoing progress monitoring of teacher classroom practices that can inform data-based decision-making on how best to tailor interventions and supports to meet teachers' professional development needs. A multi-tiered formative assessment approach to teacher evaluation provides teachers with a mechanism for receiving timely, actionable, and data-driven feedback to enhance their practice and promote positive student learning outcomes (Reddy et al. 2013c). An RTI system for teachers would include the same foundational principles but additionally integrating the focus on teachers. The system would include the following: (a) an evidence-based core curriculum of effective teaching practices, (b) universal screening through classroom observations and other measures of teacher practice, (c) a multi-tiered system of evidence-based interventions matched to teacher needs, (d) data-based decision-making guiding professional development, and (e) progress monitoring of teacher classroom practice through formative and summative assessments.

This chapter describes a multi-tiered teacher formative assessment model embedded within an RTI framework and uses the Classroom Strategies Scale (CSS) as an illustration of one promising multidimensional measure of instructional and behavioral management practices designed to monitor teacher progress in meeting instructional goals. The theory, research, and evidence underlying the CSS are outlined in addition to its application in a case example. Finally, directions for future research are described.

Multi-Tiered Teacher Formative Assessment Model

A three-tiered model that embeds formative assessment at each tier can be utilized to enhance teacher instructional practices and provide additional support to teachers requiring more intensive or different support to improve their instruction and behavior management skills. Multi-tiered formative assessment of teachers provides a systematic process for examining the extent to which teachers are benefitting from instructional support and monitoring adaptations to the curriculum or instruction to support effective instruction and student learning. Within a public health framework, we hypothesize that tier 1 assessment and support targets all educators rather than those at risk for struggling with instruction and is designed to prevent or reduce new behavioral problems and/or academic failures from emerging in classrooms (Epstein et al. 2008; Glover and DiPerna 2007). Thus, tier 1 formative assessment would focus on routine progress monitoring to promote teachers use of effective instructional and classroom management practices. Tier 1 formative assessment could also include the provision of high-quality instructional feedback that is integrated into routine teacher evaluation systems to promote effective instruction or changes in school-wide policies to prevent behavior problems and promote effective instruction. Within this three-tiered model, we would expect the majority of teachers to effectively utilize tier 1 formative assessment data to enhance their instruction and meet their instructional goals. However, for a subset of teachers, tier 1 formative assessment will be necessary but insufficient to promote enhanced classroom practices. This may reflect level of experience, composition of students in the classroom, training, or other idiographic contextual variables that impact teacher practices (Shernoff et al. 2011a, b, 2015).

Tier 2 assessment and support would be designed for teachers at risk for struggling with behavior management or effective instruction and/or would target those teachers who make insufficient progress at the first tier. Tier 2 supports for teachers may include individualizing or small-group professional development, including being paired with colleagues who are skilled in implementing specific evidence-based interventions designed to improve teachers' behavior or rate of learning a new skill. Mentor colleagues could use the CSS to monitor their peers' implementation of specific behaviors and skills, as well as provide targeted feedback to facilitate implementation. Tier 2 supports may also target teachers who are most at risk for being socially isolated within the school because they are new (Shernoff et al. 2011b) and would benefit from being linked with experienced colleagues who are well positioned to support their professional development.

Tier 3 assessment and support would reflect increasing the duration or intensity of support to teachers who continue to struggle in meeting professional development goals. Visual performance feedback guided by CSS scores can be used to enhance teachers' use of evidence-based instructional and behavior management practices. Tier 3 support may include job-embedded instructional coaches who provide in vivo modeling, demonstration, and guided practice for those teachers needing more intensive classroom-based support (Joyce and Showers 2002; Scott et al. 2012; Shernoff et al. 2011b, 2015). Coaching can embed a problem-solving process that focuses on establishing mutually agreed upon and practical intervention plans to improve educators' classroom practices and student achievement. Although tier 3 supports are more expensive, they may be time limited and/or reserved only for those teachers who have not responded to tier 1 and tier 2 support.

The CSS: A Multi-Tiered Teacher Formative Assessment

The CSS is a promising multidimensional measure of instructional and behavioral management practices designed to monitor teacher progress in meeting instructional goals. The theoretical and empirical basis for the CSS is described in detail below.

Theoretical Foundation of the CSS

The CSS was designed for school personnel to assess educators' use of evidence-based instructional and behavioral management strategies (BMS) associated with positive student outcomes. The CSS includes two forms, an Observer Form, which is used by school personnel to evaluate teachers during classroom observations and the Teacher Form, which is used by teachers to self-reflect on their strategy usage following direct classroom observations. Both the Observer and Teacher Forms include similar information designed to: (a) enhance communication of evidence-based instructional and behavioral management practices through complimentary perspectives and (b) identify specific practice goals linked to professional development and supports.

The CSS-Observer and Teacher Form constructs and strategies are grounded in more than 60 years of effective teaching and behavioral management literature (e.g., Brophy and Good 1986; Gage 1978; Marzano 1998; Marzano et al. 2001; Wittrock 1986; Walberg 1986). This body of work highlights the general features of effective instructional practice linked to positive student academic performance (e.g., Bennet 1988; Creemers 1994; Good and Brophy 1980; Harris 1998; Hattie et al. 1996; Scheerens 1992; Walberg 1986; Wang 1991; Wang et al. 1993). The CSS scales and strategies are based on direct instruction, differentiated instruction, and constructivist models of teaching. The CSS includes two domains (instruction and behavior management) that encompass nine (subscales) dimensions related to positive student learning (see Table 1).

The CSS was developed based on several key theoretical assumptions. First, both instructional and BMS are used dynamically to enhance classroom climate and student learning. Second, teachers use *sets of strategies* both in concert and in series to each other to foster student learning. Therefore, supports for implementing the evidence-based strategies should target sets of teaching strategies and the sequence of strategies to use in instruction. Third, the CSS dimensions (subscales) are assumed to be correlated based on prior research and theory that underscores that teaching is an interactive process that requires teachers to fluidly adapt strategies as lessons proceed (Clark and Peterson 1986; Harris 1998; Tomlinson and McTighe 2006). Fourth,

Table 1 Definitions of the three-part CSS assessment. Table 1 is published Reddy and Dudek (2014)

Part 1 Strategy Counts[a]	*Definitions*
Concept summaries	A teacher summarizes or highlights key concepts or facts taught during the lesson. Summarization statements are typically brief and clear. This teaching strategy helps students organize and recall material taught
Academic response opportunities	A teacher creates opportunities for students to share their understanding of the lesson content with the teacher or class. These opportunities can be verbal or nonverbal responses (e.g., explain answers, repeat key points, brainstorm ideas, and show answers on the board)
Academic praise statements	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for appropriate academic performance
Academic corrective feedback	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for incorrect academic performance
Clear one- to two-step directives	A teacher gives a verbal instruction that specifically directs a behavior to occur immediately. These directives are clear and they provide specific instructions to students to perform a behavior. They are declarative statements (not questions), describe the desired behavior, and include no more than two steps
Vague directives	A teacher gives a verbal instruction that is unclear when directing a behavior to occur immediately. These directives are vague, may be issued as questions, and often include unnecessary verbalizations or more than two steps
Behavioral praise statements	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for appropriate behavior
Behavioral corrective feedback	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for inappropriate behavior
Total	The sum of the frequency of the eight teacher behaviors
Part 2 Instructional Strategies Scales	*Definitions*
Total scale	The Total Instructional Strategies scale reflects the overall use of instructional methods and academic monitoring/feedback
Instructional methods; composite scale	How classroom instruction occurs. Measures teachers' use of teacher directed student directed methods, or differentiated instruction. This includes how a teacher incorporates active learning techniques such as hands on learning and collaborative learning in the presentation of lessons as well as how a teacher delivers academic content to students
Adaptive instruction; subscale	Strategies teachers use to respond to their students' learning needs while teaching. These practices reflect teacher flexibility and responsiveness to students' needs, as well as methods of differentiated instruction
Student-directed instruction subscale	Strategies teachers use to actively engage students in the learning process. These practices encompass constructivist and hands-on instructional techniques, linking lesson content to prior learning, personal experiences, and cooperative learning
Direct instruction; subscale	Strategies teachers use to deliver academic content or convey information to students. These practices include direct instruction techniques, modeling, identifying, and summarizing
Academic monitoring/feedback; composite scale	How teachers monitor students' understanding of the material and provide feedback on their understanding. These strategies assess students' thinking and encourage students to examine their own thought processes. Teachers guide students understanding by encouraging students, affirming appropriate application of the material, and correcting misperceptions
Promotes students' thinking; subscale	Strategies teachers use to activate students' thinking about the lesson material. These practices assess teachers' efforts to get their students to think about their thinking process (i.e., open-ended, what, how, and why)
Academic performance feedback; subscale	Strategies teachers use to provide specific feedback to their students' on their understanding of the material. These practices assess teachers' efforts to explain what is correct or incorrect with student academic performance

Table 1 (continued)

Part 2 Behavioral Management Strategies Scales	*Definitions*
Total scale	The total behavioral management strategies scale reflects the overall use of proactive methods and behavior feedback
Preventative methods; composite scale	Strategies teachers use to promote positive behaviors in the classroom and reduce the likelihood of negative behaviors. These strategies include prompts, routines, reviewing rules, and presenting instructions or requests in a clear manner
Proactive methods; subscale	Verbal and nonverbal strategies teachers use to prevent student disengagement, and problem behaviors from occurring in the classroom. These practices assess how teachers create a positive classroom environment
Directives; subscale	Strategies teachers use for issuing directions or instructions to students and behavioral expectations in the classroom
Behavioral feedback; composite scale	How teachers respond to students appropriate and inappropriate behaviors. This includes the usage of praise to encourage positive behaviors and corrective feedback to redirect negative behaviors
Praise subscale	Verbal and nonverbal strategies teachers use to positively reinforce specific appropriate behaviors in the classroom. These practices assess how teachers respond to positive behavior in the classroom
Corrective feedback; subscale	Verbal and nonverbal strategies teachers use to correct students' inappropriate behavior. These practices assess how teachers respond to negative behavior in the classroom
Part 3 Classroom Checklist Items	
1. Different methods/mediums of instruction are present in the classroom (e.g., blackboard, overhead projector, smart board, student clickers)	
2. Learning aids are present in the classroom (e.g., number chart, vocabulary list, critical thinking questions)	
3. Learning materials are present in the classroom (e.g., pencils, rulers, construction paper)	
4. Learning materials and areas in the classroom are labeled	
5. A procedure or routine exists for students to organize their desks, backpacks, or learning materials	
6. Classroom (e.g., floors, walls, table) is clean and uncluttered	
7. Tables/desks are arranged for students to easily view and participate in the lesson	
8. Classroom lesson or activity schedules are clearly posted	
9. Assignments (e.g., homework, readings, tests) are clearly posted	
10. Student work, artwork, and accomplishments are displayed in the classroom	
11. Methods for tracking student academic and/or behavioral progress (e.g., homework-tracking chart, rule-following chart, sticker/star chart) are present	
12. Classroom-wide reward system is present (e.g., ticket bin for a pizza party)	
13. Classroom rules are posted	
14. Classroom rules specify positive behaviors that students "should do" rather than "not do"	

[a] CSS-Teacher Form does not include Part 1 Strategy Counts—includes Parts 2 and 3
CSS Classroom Strategies Scale

sets of strategies are context dependent to lesson format and student learning needs. The literature devoted to measuring differentiated instructional approaches has documented the effectiveness of tailoring learning interventions to meet individual student needs (Ames 1992). Finally, the fifth theoretical assumption asserts effective teaching is a continuous learning process which requires ongoing specific feedback and supports that inform the planning and implementation of effective instructional and behavioral management practices. Teachers want and need timely, specific, actionable, and data-driven feedback to inform their instruction.

The CSS was developed as a user-friendly multidimensional assessment of instructional and BMS. The CSS generates scores that: (1) assess educators' use of empirically supported instructional and classroom BMS, (2) identify practice goals for improvement, (3) monitor educators' progress towards practice goals following intervention, (4) provide evidence for professional development and supports (e.g., professional learning committees), and (5) refine school-wide teacher professional development plans.

Development of the CSS

The CSS-Observer and Teacher Forms were iteratively developed over the course of 4 years using contemporary test theory (e.g., Anastasi and Urbina 1997; Benson 1998; Crocker and Algina 1986; Kane 2002, 2008). As noted, the CSS was designed specifically for use by school personnel in routine educational practice with the central goal to assess teachers' *use* of evidence-based instructional and behavioral management practices and to *inform* professional development and coaching efforts (e.g., Reddy and Dudek 2014; Reddy et al. 2013a; Reddy, Dudek, Fabiano, & Peters, 2015).

The development process was guided by several methods: (1) expert input, (2) consumer input, (3) extensive field-testing with more than 500 classrooms, and (4) a set of data analytic methods. Domains and strategy/items were developed through input from a national advisory board of experts in instruction, behavior management, and measurement, and a comprehensive review of peer-reviewed publications and other related tests. Consumer input (advisory boards of principals, general, and special education teachers) provided important feedback on the specific domains and strategies/items, as well as item ambiguity and possible bias. Face/content validity of the CSS was established in part through the expert and consumer advisory boards (see Reddy et al. 2013a, 2015 for a detailed description). The boards were also encouraged to provide feedback on new domains and strategies/items and the CSS intended use and score utility for assessing practices and informing changes practices (i.e., professional development). Additionally, numerous data analytic methods were employed to refine the CSS scales and strategies/items such as item to total correlations, pooled mean item variances across observation (level of disagreement), as well as confirmatory factor analysis within observation using recommended fit indices (Jackson et al. 2009) and information-theory-based indices of relative fit (Bowen and Guo 2012; see Reddy et al. 2013a; 2013b and Reddy, Dudek, Fabiano & Peters, 2015 for details).

The CSS Model

The CSS-Observer Form consists of three parts (see Table 1). For Part 1 Strategy Counts, observers tally each time eight instructional and behavior management strategies are used by the teacher during a classroom observation (lesson) and whether the strategy used was for individual students or group of students (i.e., two or more students). For Part 2 Strategy Rating Scales, observers complete the instructional scale (IS) and BMS scale after classroom observations. After the completion of the Part 2 Strategy Rating Scales, observers complete Part 3 the Classroom Checklist which includes 14 items that assess the presence of classroom structural procedures. The CSS-Teacher Form does not include the Part 1 Strategy Counts, but contains the exact same Part 2 Strategy Rating Scales items for IS and BMS, as well as the same Part 3 Classroom Checklist items.

Part 1 Strategy Counts—Observer Form Strategy Counts includes eight strategies that are nested in the domains of instructional and behavioral management practices. Concept summaries, academic response opportunities, academic praise, and academic corrective feedback fall under instruction, while one- to two- step directives, vague directives, behavioral praise, and behavioral corrective feedback fall under behavior management.

Concept summaries are defined as the teacher summarizing or highlighting key concepts or facts (steps) taught throughout the lesson (not at the beginning or end of lessons). Examples may include "we learned today that a hypothesis is a scientist's best guess about how an experiment will turn out" and "DeShawn, to find a cube's volume we multiply the length times the width, times the height." The use of concept summaries has been linked to enhancing students' understanding of the lesson and recall and organization of learning material (e.g., Brophy and Alleman 1991; Brophy 1998; Hines et al. 1985; Rosenshine and Stevens 1986). Techniques that reinforce key concepts and facts (concept summaries) are important for students with executive functioning impairments and/or auditory processing disorders as these students often require lesson content to be repeated and emphasized for successful integration and skill application (Reddy et al. 2015). A general guideline for educators is to provide brief concept summaries every 2–3 min during learning activities.

Academic response opportunities are an important strategy for teachers to encourage students to share their ideas and understanding of lesson content in class. Academic response opportunities can be verbal or nonverbal responses (e.g., explain answers, repeat key points, brainstorm ideas, and show answers on the board). Research on "opportunities to respond" (OTR) among students with behavior disorders underscores the utility of teachers creating opportunities for their students to respond to questions and learning activities during instructional time (e.g., Partin et al. 2010; Sutherland et al. 2003; Stichter et al. 2009). Research has found that increasing OTR and praising students for effort and/or correct responses can lead to higher levels of both on task behavior, prosocial behavior, and correct student answers (Sutherland et al. 2002). The recommendation is that teachers elicit 4–6 responses per minute when teaching new learning material and 9–12 responses per minute when reviewing previously taught learning material (practice or drill work).

Academic feedback (i.e., praise and corrective feedback) is important for instruction and student learning (e.g., Council for Exceptional Children; Bender 2008; Gable et al. 2009; Tomlinson and Edison 2003). Praise for academic performance is verbal or nonverbal statements or gestures provided by teachers to individual or groups of students immediately following academic responses. Likewise, corrective feedback for academic performance is verbal or nonverbal statements or gestures provided by teachers to individual or groups of students immediately following incorrect academic responses.

Praise statements should be frequent, immediate, enthusiastic, and specifically describe the behaviors of the student's success. Praise statements that are implemented consistently will orient students towards a better appreciation of students' own task-related behavior (Brophy 1981; Gable et al. 2009). Similarly, academic corrective feedback for students' incorrect academic responses should be immediate and explain what is specifically incorrect about their answers (Hattie and Timperely 2007). Simply telling children their answer is right or wrong is not enough. Students need to be told what specifically is correct or incorrect about their answers (e.g., "Your answer is incorrect because you forgot to carry over the one when you added the tens column"). A common sequence of instructional strategies is to first provide an academic response opportunity followed by a praise statement or corrective feedback for academic performance. For all students, academic response opportunities and academic feedback (praise and correct feedback) are critically important for encouraging and monitoring task completion.

Effective directives are important for enhancing appropriate behavior and follow through on tasks (Kern and Clemens 2007). Clear one- to two-step directives are brief verbal instructions that direct specific student behavior. Directives are clear, declarative statements (not questions or favors) that specifically describe the desired behavior in no more than two steps. Directives are most effective when phrased as "do" commands (i.e., telling children what they should do rather then what they should not). On the other hand, vague directives are verbal directives that are unclear, issued as questions, and include unnecessary verbalizations (more than two steps).

For the CSS, behavioral feedback consists of two strategies: praise for appropriate behavior and corrective feedback for inappropriate behavior. Praise for appropriate behavior is verbal or nonverbal statements or gestures provided by teachers to individuals or groups of students immediately following appropriate behavior.

As with praise for academic performance, key features of effective praise statements for behavior are high frequency, immediacy, and specifically labeling the appropriate behavior. Research has shown that levels of on task behavior significantly increase when students were given specific praise about their behavior compared to simply positive praise (i.e., Brophy 1981; Chalk and Bizo 2004). In contrast, corrective feedback for inappropriate behavior is verbal or nonverbal statements or gestures provided by teachers to redirect inappropriate behavior. Like praise, corrective feedback should be specifically labeled and given after inappropriate behavior is observed (Bangert-Down et al. 1991). Research has recommended that teachers should provide approximately *one* corrective feedback statement for *every three* praise statements in the classroom (Stitcher et al. 2009; White 1975).

Part 2 Strategy Rating Scales and Part 3 Classroom Checklist—Observer and Teacher Form Following classroom observations, observers (teachers) complete the Part 2 Strategy Rating Scales, IS and BMS scales (see Table 1). The IS scale includes 28 items that comprise a total scale, two composite scales, and five subscales. The Instructional Methods Composite Scale (17 items; maximum frequency score of 119) consists of the Direct Instruction (8 items; maximum score of 56), Adaptive Instruction (4 items; maximum score of 20) and Student Focused Instruction (6 items; maximum score of 42) subscales. The Academic Monitoring/Feedback Composite scale (11 items; maximum score of 77) consists of the Promotes Student Thinking (5 items; maximum score of 35) and Academic Performance Feedback (6 items; maximum score of 42) subscales.

The BMS scale includes 26 items that compose a total scale, two composite scales, and four subscales. The Behavioral Feedback Composite scale (12 items; maximum frequency score of 84) consists of Praise (5 items; maximum score of 35) and Corrective Feedback (7 items; maximum score of 49) subscales. The Proactive Methods Composite scale (14 items; maximum score of 91) consists of Prevention Management (8 items; maximum score of 56) and Directives (6 items; maximum score of 42) subscales.

After classroom observations, observers (teachers) rate how often (*Frequency Rating*) teachers used specific instructional and BMS on a 7-point Likert scale (1 "never used," 4 "sometimes used," 7 "always used") and then rate how often the teachers *should have* used each strategy (*Recommended Frequency)* on a 7-point Likert scale (1 "never used," 4 "sometimes used," 7 "alwaysused"). The Part 2 Rating Scales produce both frequency scores and discrepancy scores. For the Part 2 Strategy Rating Scales, item discrepancy scores are computed as follows: **|** recommended frequency—frequency ratings **|.** In sum, absolute value discrepancy scores indicate if any change (regardless of direction) was needed as measured by the observer (teacher) using the CSS. Larger discrepancy score values indicate greater amounts of change are needed in the practices measured by the CSS (see Reddy et al. 2013a for details on scoring). After completing Parts 1 and 2, the observer (teachers) completes the Classroom Checklist (Part 3). The Classroom Checklist assesses the presence of 14 specific items or procedures in the classroom (see Table 2).

Training

The CSS-Observer Form training consists of a five-step process focused on content knowledge and direct observational skills for the specific components of the CSS. Training includes: (1) group training, (2) knowledge testing, (3) practice coding, (4) criterion testing, and (5) report interpreting, coaching training.

Training includes:

(1) Didactic group trainings from a CSS Trainer/Master Coder which include discussions of definitions, criteria, and orientation to the scientific literature guiding the development of the CSS and the recommended frequencies of Part 2 strategies.
(2) Completion of knowledge tests for CSS Part 1–3.
(3) Practice coding classroom videos using the CSS and review of practice results by a CSS Trainer/Master Coder.

Table 2 Using teacher formative assessments: implications for practice

RTI components	Teacher support	Organizational support
Tier 1 formative assessment	Routine progress monitoring to promote continuous learning opportunities Promote reflection, dialogue, and ongoing PD Increase use of general evidence-based instructional practices Increase use of best practices for teacher assessment linked to PD	Inform district-wide professional development (PD) needs and goals Link individual performance with school-wide and district-level performance goals Inform school-wide policies to prevent behavior problems Inform school-wide policies to enhance instructional goals Promote organizational culture that supports ongoing learning and continued PD
Tier 2 formative assessment	Enhance relationships and supports for at risk teachers Mentor at-risk teachers via indigenous school supports Increase use of targeted evidence-based instructional practices Increase use of targeted practices for teacher assessment linked to PD	Promote workforce development for teachers at risk Identify organizational goals to support teachers at risk for struggling with behavior management or effective instruction Promote sustainability of teacher support and PD
Tier 3 formative assessment	Provide intensive individualized PD for identified teachers Provide intensive individualized teacher assessment linked to PD for identified teachers	Enhance communication between leadership teams and teachers regarding PD goals and benchmarks Detail action plans to enhance school-wide knowledge, skill, and learning Facilitate the link between classroom/individual learning and school-wide/organizational growth

(4) Completion of five classroom video coding criterion tests using the CSS. Certification as a qualified user is set by reaching the minimum inter-rater reliability level of 80% with CSS Trainer/Master Coders.
(5) CSS score report interpretation and coaching.

Psychometric Characteristics

Construct Validity For the CSS Parts 1 through 3, the constructs and strategies/item were extensively developed through ongoing expert and consumer advisory board feedback, pilot testing, and advance statistical analyses. All three parts represent important instructional and BMS associated with effective teaching linked to general and special education student academic and behavioral outcomes (e.g., Reddy and Dudek 2014; Reddy et al. 2012; Reddy et al. 2015a; 2015b).

The Part 2 IS and BMS scales are theoretically and factor analytically derived (confirmatory factor analysis) within classroom observations. The CSS factor structure was examined with more than 12 confirmatory factor analyses using generalized least squares estimation (SPSS's AMOS Version 19 software, Arbuckle 2010). As described in Reddy et al. (2013a), several fit indices (e.g., χ^2/d*f,* root mean square error of Approximation (RMSEA), adjusted goodness of fit index (AGFI), and goodness of fit index (GFI) recommended by Jackson et al. (2009) were used to test the fit to the data. As noted, the CSS IS and BMS subscales were conceptualized as being correlated factors based on theory and research that teachers use a blend of strategies and sets of strategies in the teaching and learning process. In addition, CSS first-order four-factor models (IS and BMS preferred factor models) were compared to alternative models including five- and six-factor models using information-theory-based indices of relative fit (i.e., Akaike Information Criterion (AIC), Browne-Cudeck Criterion (BCC), Schwarz Bayesian Information Criterion (BIC)) described by Bowen and Guo (2012). Overall, results indicated that CSS four, five- and six-factor models yielded good fit to the data and superior fit to the data in comparison to alternative models using information-theory-based indices of relative fit (see Reddy et al. 2013a). The CSS-Teacher Form also yielded good fit to the data and superior fit to the data in comparison to alternative higher order factor models (Reddy, Dudek, Fabiano & Peters, 2015).

Reliability and Validity In the high-stakes era of teacher evaluation, evidence of reliability and validity for new teacher measures is critically important for evaluating teacher effectiveness and informing professional development/coaching models. The reliability of the CSS has been examined in four studies that assess the internal consistency, inter-rater reliability, test retest reliability, and freedom from item bias.

The CSS was found to demonstrate good internal consistency (Cronbach alphas of 0.92–0.93) across Parts 1 through 3 (see Reddy et al. 2013a for further details). Inter-rater reliability data was found to have good inter-rater reliability estimates for Parts 1–3. For example, inter-rater reliability for the Part 1 Strategy Counts was $r=0.94$ (percent agreement 92 %), Part 2 IS and BMS Strategy Rating Scales was $r=0.80$ and $r=0.72$ (percent agreement 92 % and 88 %), and Part 3 Classroom Checklist was $r=0.86$ (percent agreement 91 %). The CSS inter-rater reliability estimates are consistent with accepted values for other classroom observation assessments such as the measures used in the Measures of Effective Teacher Project (Cantrell & Kane, 2013; Kane and Staiger 2012) and the Classroom Assessment Scoring System (CLASS; Pianta et al. 2008). Furthermore, the CSS has been found to have fair to good test–retest reliability (approximately 2–3 weeks). For example, an r of 0.70 (percent agreement 81 %) was found for the Part 1 Total Behaviors, *rs* of 0.86 and 0.80 (percent agreement 93 and 85 %) for the Stage 2 IS and BMS Total scales, and r of 0.77 (percent agreement was 81 %) for the Stage 3 Classroom Checklist**.** Part 2 Strategy Rating Scales items evidenced freedom of item bias for teacher age, educational degree, and years of teaching experience using differential item functioning analyses (Reddy et al. 2013a). Similar results have been found for the CSS-Teacher Form (Reddy, Dudek, Fabiano & Peters, 2015).

Validity The CSS-Observer Form evidences convergent and discriminant validity, as well as predictive validity. The CSS was compared to the CLASS, a well-established measure of teacher and classroom quality (Pianta et al. 2008). As hypothesized, the CSS corresponded with logically related CLASS domains (e.g., Behavior Management) and it did not correspond with domains hypothesized to be unrelated (e.g., Language Modeling), suggesting that the CSS yields good convergent and discriminant validity with the CLASS (Reddy et al. 2013). Results highlight the unique features of the CSS for assessing teacher classroom practices. Similar results have been found with the CSS-Teacher Form (Reddy & Dudek, under review).

The CSS also evidences good predictive validity with student academic outcomes. Using a series of two-level hierarchical linear modeling, the CSS IS scale discrepancy scores uniformly predicted student mathematics and language arts statewide testing scores for 663 third, fourth, and fifth graders (Reddy et al. 2013b). Findings revealed that IS scale discrepancy scores significantly predicted mathematics *and* English language arts proficiency scores: Relatively larger discrepancies on observer ratings of what teachers did versus what should have been done were associated with lower proficiency scores. In sum, the greater the need for changes in practices exhibited by a teacher, the greater number of students receiving lower proficiency scores.

For mathematics statewide assessments, odds ratios indicated that an *increase* in 1 standard deviation (SD) of IS discrepancy scores (i.e., Total, Instructional Methods and Academic Monitor/Feedback Composite Scales) *reduced* the odds of success (i.e., students scoring proficient or above proficiency) by approximately 25 %. Likewise, for English language arts statewide assessments, odd ratios revealed that an *increase* in 1 SD of IS discrepancy scores (i.e., Total, Instructional Methods Composite) *reduced* the odds of success by approximately 30 %. Results support the predictive validity of the CSS and have important implications for professional development. Preliminary results offer support for the predictive validity of the CSS-Teacher Form (Reddy & Dudek, under review).

CSS Case Example

The following case example demonstrates the utility of embedding teacher formative assessment within a tiered model to support teacher development and enhance instructional practices (see Table 2). The CSS provides an illustration of a progress-monitoring instrument that supports and promotes data-based decision-making regarding instructional practices and professional performance.

John is a 41-year-old, Caucasian male teacher with a bachelor's degree in elementary education. He has 13 years of experience as a teacher in elementary school settings and has worked in his current position as a fifth-grade teacher for the past 10 years. He has predominantly taught the language arts component of the fifth-grade curriculum at his district while his partner teacher instructs students in the mathematics, science, and social studies components. John teaches to one group of students in the morning (9 a.m. to 11:30 a.m.) and to a second group of students in the afternoon (12:30 p.m. to 3 p.m.). John and his partner teacher switch classrooms after recess. John's afternoon classroom is composed of 25 general education students. Although no students are classified with a specific learning disability, John noted there are six students evidencing academic and behavioral learning difficulties.

John's school employs an RTI model for its students and recognizes how this model can support teachers. To help promote effective instruction in the classroom and support teacher development, the school has recently incorporated this model into its comprehensive human capital management system (HCMS; Odden & Kelly, 2008). In its first tier, the system relies on routine educator evaluation for providing feedback to teachers on their instructional practices. This year, John will receive a series of four observations, one per quarter, by a qualified evaluator utilizing the CSS.

For his first observation in October, John and the evaluator complete the CSS. John utilizes the CSS-Teacher Form and the evaluator observes John using the CSS-Observer Form. Following the observation, John and the evaluator discuss their findings. On the CSS-Teacher Form, John indicated that he utilized behavioral corrective feedback very frequently and thought there was significant need to use it at a high frequency due to the frequency of inappropriate behaviors in the classroom. The evaluator agreed with John's assessment but also noted on the CSS-Observer Form that John rarely used proactive methods of managing students' behavior and used behavioral praise at a low frequency. The evaluator discussed with John the benefits of using these strategies to help shape a positive learning environment and reduce inappropriate behaviors in the classroom. The evaluator and John devise an operationalized behavioral plan for implementing these strategies in the classroom and set a target goal for John's next evaluation in January.

For the second observation in January, the same procedure is repeated by John and the evaluator. The results were highly similar to the first observation: frequent behavioral corrective feedback but minimal use of proactive strategies and behavioral praise. During the post-observation conference, John and the evaluator discuss the target goals established for this observation. Although John attempted to utilize the strategies recommended in the behavioral plan, John agreed with the evaluator that these were unsuccessful and that he needs additional support.

As part of the second tier of the HCMS, the school district has a collaborative teaching mentor program that pairs both new and struggling teachers with an experienced teacher that excels at using evidence-based strategies. The evaluator helps John select a teacher that exemplifies effective classroom BMS. Over the course of several weeks, the mentor teacher observes John's classroom and uses the CSS to gather information. John and the mentor teacher meet collaboratively after school several times to discuss effective behavior management strategies and to specifically work on proactive strategies and behavioral praise. For each meeting, the mentor teacher graphs John's use of BMS on the CSS to help John monitor his progress in practicing and implementing these strategies. After several weeks, the graphs show

John increasing his use of effective proactive strategies and behavioral praise. The graphs also show decreased use of behavioral corrective feedback.

For the third observation that year, both John and the evaluator notice improvement on the CSS. Through increased use of proactive strategies and behavioral praise, John was able to create a better learning environment in his classroom and reduce the amount of inappropriate student behaviors. The evaluator congratulates John on his efforts and encourages him to continue implementing these evidence-based strategies. For his fourth observation, John received very positive reviews, especially in the area of behavior management.

Had efforts at the second tier of intervention not worked, John would have been assigned more intensive interventions and supports associated with the third tier of intervention. The third tier of interventions might have included working with a professional development coach external to the school district or enrolling John in professional development courses for effective teaching. Additionally, more intensive procedures like in vivo modeling or shared teaching might also have been used to help facilitate John's competencies in these skills.

Routine progress monitoring in John school's district allowed for early identification and remediation in John's behavioral management practices. Had a more traditional evaluation system model been used, which historically requires one observation per year for tenured teachers, John's difficulties in behavior management competencies might not have been noticed until it was too late for remediation to yield gains in the classroom (Toch and Rothman 2008; Weisberg et al. 2009). Subsequently, increasing the level of support required by John allowed for tailored and specific professional development to occur. Ultimately, by implementing a multi-tiered formative assessment model for teacher development, both teachers and their students will benefit from the enhanced instruction in the classroom.

Future Research and Practice

Given that multi-tiered teacher formative assessment models are slowly emerging, several areas for future research and practice are needed to better understand the impact of tiered assessment and support for teachers (Table 2). For example, further studies must establish reliable and valid decision-making criteria for determining progression through the three tiers and what constitutes adequate progress within an RTI assessment and intervention model for teacher professional development (Glover and DiPerna 2007). Given the prominent role that formative assessment models play in our educational system, further research in this area has implications for assessing student progress over time as well as teachers. Although, a significant potential benefit of an RTI framework is its utility in determining teacher responsiveness to support, it is unclear what the infrastructure of support should be for those teachers who continue to struggle despite intensive intervention across tiers. Given a corpus of research supports the importance of high-quality implementation of evidence-based practices, more research is needed to establish methods and measures that maximize the uptake, maintenance, and implementation of evidence-based interventions in schools (Becker et al. 2013). Several gaps in the coaching literature warrant further attention, including a better understanding of the functional components and key elements of coaching (e.g., Blachowicz et al. 2005; Kretlow and Bartholomew 2010) and the key instructional elements (e.g., modeling, demonstration, observation) utilized (Shernoff et al. 2015). In addition, despite the important role that coaching can play in supporting teachers in need of tier 3 support, existing studies provide very limited insight into the feasibility of coach-based professional development and the extent to which such models can be sustained over time (Becker et al. 2013; Nadeem et al. 2013; Shernoff et al. 2015).

Conclusion

Multi-tiered teacher formative assessment models are a viable alternative to traditional, high stakes, summative assessment of teachers and are aligned with RTI models being adopted in the US. This chapter underscores the importance of teacher assessment, ongoing feedback, and providing a continuum of support for teachers in order to maximize teacher professional development and student learning. The CSS was illustrated as one promising tool that could be used to develop meaningful evidence-based feedback for teachers and improve the quality of implementation of evidence-based interventions used in their classrooms. Coaching was described as one mechanism for providing more intensive, job-embedded support to those teachers requiring more intensive support.

References

Ames, C. (1992). Classrooms: Goals, structures and student motivation. *Journal of Educational Psychology, 84,* 261–271.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River: Prentice Hall.

Arbuckle, J. L. (2010). *IBM SPSS Amos 19 User's Guide* (pp. 587–590). Crawfordville: Amos Development Corporation.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213–238.

Barnett, W. S. (2004). *Better teachers, better preschools: Student achievement linked to teacher qualifications*. http://nieer.org/resources/policybriefs/2.pdf. Accessed 30 Nov 2013.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to intervention policy considerations and implementation*. Reston: National Association of State Directors of Special Education.

Batsche, G., Curtis, J., Dorman, C., Castillo, J., & Porter, L. J. (2008). The Florida problem solving/response to intervention model: Implementing a statewide initiative. In S. Jimerson, M. Burns, & A. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 378–395). New York: Springer.

Becker, K. D., Bradshaw, C. P., Domitrovich, C., & Ialongo, N. S. (2013). Coaching teachers to improve implementation of the good behavior game. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(6), 482–493.

Bender, W. N. (2008). *Differentiating instruction for students with learning disabilities: Best teaching practices for general and special educators* (2nd ed.). Thousand Oaks: Corwin Press.

Bender, W. N., & Shores, C. (2007). *Response to intervention: A practical guide for every teacher*. Thousand Oaks: Corwin Press.

Bennet, N. (1988). The effective primary school teacher: The search for a theory of pedagogy. *Teaching and Teacher Education, 4,* 19–30.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 1*(7), 10–22.

Blachowicz, C. L., Obrochta, C., & Fogelberg, E. (2005). Literacy coaching for change. *Educational Leadership, 62*(6), 55.

Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. New York: Oxford University Press.

Brophy, J. E. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*(1), 5–32.

Brophy, J. (1998). *Motivating students to learn*. New York: McGraw-Hill.

Brophy, J., & Alleman, J. (1991). A caveat: Curriculum integration isn't always a good idea. *Educational, Leadership, 49*(2), 66.

Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research in teaching* (3rd ed., pp. 328–375). New York: Macmillian.

Cantrell, S., & Kane, T. J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf. Accessed 11 Nov 2013.

Carlisle, J. F., Correnti, R., Phelps, G., & Zeng, J. (2009). Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading. *Reading and Writing: An Interdisciplinary Journal, 22,* 459–486.

Chalk, K., & Bizo, L. A. (2004). Specific praise improves on-task behavior and numeracy enjoyment: A study of 4 pupils engaged in the numeracy hour. *Educational psychology in Practice, 20,* 335–351.

Chunn, M., & Witt, J. (2008). *System to enhance educational performance*. http://www.isteep.com/index.html. Accessed 25 Nov 2013.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Connor, C. M., Son, S. H., Hindman, A., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, and family characteristics: Complex effects on first graders' language and early reading. *Journal of School Psychology, 43,* 343–375.

Crawford, A. D., Zucker, T. A., Williams, J. M., Bhavsar, V., & Landry, S. H. (in press). Initial validation of the pre-kindergarten classroom observational tool and goal setting system for data-based coaching. *School Psychology Quarterly*.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional learning*. Princeton: Educational Testing Service.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis and Archives, 8*(1), 1–44.

Epstein, J. L., Sanders, M. G., Sheldon, S. B., Simon, B. S., Salinas, K. C., Jansorn, N. R., Van Voorhis, F. L., Martin, C. S., Thomas, B. G., Greenfeld, M. D., Hutchins, D. J., & Williams, K. J. (2008). *School, family, and community partnerships: Youth handbook for action* (3rd ed.). Thousand Oaks: Corwin Press.

Espinosa, L. M. (2002). *Preschool policy matters; High-quality preschool: Why we need it and what it looks like*. http://nieer.org/resources/policybriefs/1.pdf. Accessed 30 Nov 2013.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93–99.

Gable, R., Hester, P., Rock, M., & Hughes, K. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Intervention in School & Clinic,* 44(4), 195.

Gage, N. L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.

Gersten, R., Chard, D. J., & Baker, S. (2000). Factors enhancing sustained use of research-based instructional practices. *Journal of Learning Disabilities, 33,* 453.

Glover, T. A., & DiPerna, J. C. (2007). Service delivery for Response to Intervention: Core components and directions for future research. *School Psychology Review, 36,* 526–542.

Good, T. L., & Brophy, J. (1980). *Educational psychology: A realistic approach* (2nd ed.). New York: Holt, Rinehart, & Winston.

Griffiths, A., Burns, M. K., Parson, L. B., Tilly, W. D., & VanDerHeyden, A. (2007). *Response to intervention: Research for practice.* Alexandria: National Association of State Directors of Special Education.

Harris, A. (1998). Effective teaching: A review of the literature. *School Leadership & Management, 18,* 169–183.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, (1), 81–112.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on students learning: A meta-analysis. *Review of Educational Research, 66*(2), 99–136.

Heneman, H. G. III, & Milanowski, A. (2004). Alignment of human resource practices and teacher performance competency. *Peabody Journal of Education, 79*(4), 108–125.

Hines, C. V., Cruickshank, D. R., & Kennedy, J. J. (1985). Teacher clarity and its relationship to student achievement and satisfaction. *American Educational Research Journal, 22,* 87–99.

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Jackson, D. L., Gillapsy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14,* 6–23.

Jackson, S., Pretti-Frontczak, K., Harsjusola-Webb, S., Grisham-Brown, J., & Romani, J. (2009). Response to intervention: Implications for early childhood professionals. *Language, Speech and Hearing Disorders in Schools, 40,* 1–11.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools*. New York: Basic Books.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria: Association for Supervision and Curriculum Development.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21,* 31–41.

Kane, M. (2008). Terminology, emphasis, and utility in validation [Peer Commentary on the paper A suggested change in terminology and emphasis regarding validity and education]. *Educational Researcher, 37*(2), 76–82.

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. MET Research Paper. Seattle, Washington: Bill & Melinda Gates Foundation. http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf. 16 July 2012.

Kern, L., & Clemmens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools, 44,* 65–75.

Klinger, J. K., & Edwards, P. A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly, 41,* 108–117.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33,* 279–299.

Leigh, A. (2010). Estimating teacher effectiveness from two-year changes in students' test scores. *Economics of Education Review, 29,* 480–488.

Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora: Mid-continent Research for Education and Learning. (Eric Document Reproduction Service No. ED 427 087).

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria: Association for Supervision and Curriculum Development.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria: ASCD.

Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research, 80*(4), 242–247.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice, 24*(4) 186–195.

Nadeem, E., Gleacher, A., & Beidas, R. S. (2013). Consultation as an implementation strategy for evidence-based practices across multiple contexts: Unpacking the black box. *Administration and Policy in Mental Health and Mental Health Services Research, 40*(6), 530–540.

National Association of State Directors of Special Education. (2007). *Policy considerations and implementation and response to intervention: NASDSE and CASE white paper on RTI*. www.nasdse.org. Accessed 10 Nov 2013.

National Center on Response to Intervention. (2010). *Essential components of RTI—A closer look at response to intervention*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education. http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf. Accessed 11 Nov 2013.

National Reading Panel. (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Research Council. (2001). *Educating children with autism.* committee on educational interventions for children with autism. In L. Catherine & J. P. McGee (Eds.), Division of behavioral and social sciences and education. Washington, DC: National Academy Press.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Odden, A., & Kelly, J. (2008). *Strategic management of human capital in public education.* Madison: University of Wisconsin, Wisconsin Center for Education Research, Consortium for Policy Research in Education, Strategic Management of Human Capital.

Partin, T., Robertson, R., Maggin, D., Oliver, R., & Wehby, J. (2010). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure, 54,* 172–178.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system [CLASS] manual: Pre-K*. Baltimore: Brookes Publishing.

Pianta, R. C., Hamre, B. K., & Downer, J. (2011). Aligning measures of quality with professional development goals and goals for children's development. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 297–315). Baltimore: Paul H. Brookes Publishing Co.

Reddy, L. A., & Dudek, C. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School and Educational Psychology, 2,* 71–84.

Reddy, L., Fabiano, G., Barbarasch, B., & Dudek, C. (2012). Behavior management of students with attention-deficit/hyperactivity disorders using teacher and student progress monitoring. In L.M. Crothers & J.B. Kolbert (Eds.), *Understanding and managing behaviors of children with psychological disorders: A reference for classroom teachers*. New York: Continuum International Publishing Group, Inc.

Reddy, L. A., Fabiano, G., Dudek, C., & Hsu, L. (2013a). Development and construct validation of the classroom strategies scale–observer form. *School Psychology Quarterly, 28,* 317–34.

Reddy, L. A., Fabiano, G., Dudek, C. M., & Hsu, L. (2013b). Predictive validity of the classroom strategies scale–observer form on statewide testing. *School Psychology Quarterly, 28,* 301–316.

Reddy, L. A., Fabiano, G., Dudek, C., & Hsu, L. (2013c). Instructional and behavior management practices implemented by elementary general education teachers. *Journal of School Psychology, 51*(6), 683–700.

Reddy, L., Fabiano, G., & Jimerson, S. (2013). Assessment of general education teachers' tier 1 classroom practices: Current science and practice. *School Psychology Quarterly, 28*(4), 273–276.

Reddy, L. A., Fabiano, G., & Dudek, C. (2013). Concurrent validity of the classroom strategies scale–observer form. *Journal of Psychoeducational Assessment, 31*(3), 258–270.

Reddy, L. A., Kettler, R. J., & Kurz, A. (2015). Schoolwide educator evaluation for improving school capacity and student achievement in high poverty schools: Year 1 of the school system improvement project. *Journal of Educational and Psychological Consultation*.

Reddy, L. A., Newman, E., & Verdesco, A. (2015a). Students with ADHD: Using self-regulated learning interventions and student and teacher formative assessment. In T. Cleary (Ed.), *Self-regulated learning interventions with at-risk populations: Academic, mental health, and contextual considerations*. Washington, DC: American Psychological Association Press.

Reddy, L. A., Newman, E. & Verdesco, A. (2015b). Attention-deficit hyperactivity disorders: Using cognitive behavioral interventions and teacher formative assessment in schools. In R. Flanagan & C. Allen (Eds.), *Cognitive behavioral therapies for children in schools*. New York: Springer.

Reddy, L. A., Dudek, C. M., Fabiano, G., & Peters, S. (2015c). *Measuring teacher self-report on classroom practices: Construct validity and reliability of the Classroom Strategies Scale – Teacher Form*. School Psychology Quarterly.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2) 247–252.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Witrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmillan.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*(4), 329–339.

Scheerens, J. (1992). *Effective schooling: Research, theory, and practice*. London: Cassell.

Schloss, P. J., & Smith, M. A. (1998). *Applied behavior analysis in the classroom*. Boston: Allyn and Bacon.

Scott, S. E., Cortina, K. S., & Carlisle, J. F. (2012). Understanding coach-based professional development in reading first: How do coaches spend their time and how do teachers perceive coaches' work? *Literacy Research and Instruction*, *51*(1), 68–85. doi:http://dx.doi.org/10.1080/19388071.2011.569845.

Shernoff, E. S., Marinez-Lora, A., Frazier, S. L., Jakobsons, L. J., Atkins, M. S., & Bonner, D. (2011a). Teachers supporting teachers in urban schools: What iterative research designs can teach us. *School Psychology Review, 40,* 465–485.

Shernoff, E. S., Mehta, T. G., Atkins, M. S., Torf, R., & Spencer, J. (2011b). A qualitative study of the sources and impact of stress among urban teachers. *School Mental Health*, *3*(2), 59–69. doi:10.1007/s12310-011-9051-z.

Shernoff, E. S., Lakind, D., Frazier, S. L., & Jakobsons, L. (2015). Coaching early career teachers in urban elementary schools: A mixed method study. *School Mental Health, 7,* 6–20. doi:10.1007/s12310-014-9136.

Shernoff, E. S., Jakobsons, L., Lakind, D., Frazier, S. L., Marinez-Lora, A., Hamre, B. K., Patel, D., Atkins, M. S., Parker Katz, M., Neal, J., & Smylie, M. (under review). Enhancing urban early career teachers' effectiveness and connectedness: A mixed method study. Manuscript submitted for publication to *American Education Research Journal*.

Stitcher, J. P., Lewis, T. J., Whittaker, T. A., Richter, M., Johnson, N., & Trussell, R. (2009). Assessing teacher use of opportunities to respond and effective classroom management strategies. *Journal of Positive Behavior Interventions, 11,* 68–81.

Sutherland, K. S., Wehby, J. H., & Yoder, P. J. (2002). Examination of the relationship between teacher praise and opportunities for students with EBD to respond to academic requests. *Journal of Emotional and Behavioral Disorders, 10*(1), 5–13.

Sutherland, K. S., Adler, N., & Gunter, P. L. (2003). The effect of varying rates of opportunities to respond to academic requests on the classroom behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 11,* 239–248.

Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.

Tomlinson, C. A., & Edison, C. C. (2003). *Differentiation in practice: A resource guide for differentiating curriculum, grades K-5*. Alexandria: Association for Supervision and Curriculum Development.

Tomlinson, C. A., & McTighe, C. A. (2006). *Integrating differentiated instruction (DI) and understanding by design (UbD)*. Alexandria: Association for Supervision and Curriculum Development.

Walberg, H. J. (1986). Synthesis of research on teaching. In M. Wittorck (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Wang, M. C. (1991). Productive teaching and instruction: Assessing the knowledge base. *Phi Delta Kappan, 71,* 470–478.

Wang, M. C., Haertel, G. D., & Walhberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63,* 249–294.

Ward, B. (1985). Teacher development: The challenge of the future. In S. Hord, S. O' Neal, & M. Smith (Eds.), *Beyond the looking glass* (pp. 283–312). Austin: The Research and Development Center for Teacher Education, The University of Texas.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. http://widgeteffect.org/. Accessed 2 June 2009.

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavioral Analysis, 8,* 367–372.

Wittrock, M. C. (1986). *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Part VIII

Effective Contemporary Models

Challenges Faced by New Implementation Sites: The Role of Culture in the Change Process

Dawn Miller and Rachel Freeman

Response to intervention (RTI) is a tiered approach for supporting all students within a school by providing the appropriate intensity of academic and behavioral support necessary for educational progress (Batsche et al. 2005). The critical practices that academic and behavioral RTI efforts share in common include: (a) evidence-based curricular and instructional practices for all students, (b) a data-based framework for decision-making, (c) use of a problem-solving process across all levels of the system, and (d) a team-based approach for leading, planning, and evaluating implementation efforts (Hawken et al. 2008). New implementation sites have a formidable task of operationalizing what the practices will mean to their respective building and/or district and must take into consideration the cultural issues that impact adoption and implementation.

Merriam-Webster (2012) defines culture as "the integrated pattern of human knowledge, belief, and behavior that depends upon the capacity for learning and transmitting knowledge to succeeding generations" or at a broader level culture is "the set of shared attitudes, values, goals, and practices that characterizes an institution or organization" (p. 1). The change process must be sensitive to how educators, staff members, students, and the greater community perceive RTI (Burns 2007).

Systems change often has an impact on the culture of the individuals implementing evidence-based practices, such as RTI. Districts and schools implementing RTI each have their own culture of implementation that include a complex interaction of individual and organizational dynamics (Lohrmann et al. 2013). The end result is to embed the values and beliefs that are reflected in RTI practices in the systems and practices at the district and building level. The purpose of this chapter is to (a) view the culture of change across phases of the implementation process and (b) to use the existing literature base on RTI and implementation research efforts to provide guidance to teams as they encounter challenges common in the early stages of implementation.

Cultural Characteristics of RTI Implementation

Fullan (2003) stated that in schools "Cultures consist of shared values and beliefs in the organization" (p. 57). Effective systems change in education means that beliefs and values associated with important evidence-based practices become embraced by each individual within a school. At the district level, embedding beliefs and values influences policy and adoption decisions. When considering the challenges that new RTI implementers encounter, it is believed that one must understand, first, the challenge of how culture

D. Miller (✉)
University of Kansas, Lawrence, USA
e-mail: dawnmiller@ku.edu

R. Freeman
University of Minneosota, Minneapolis, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_35

changes within any implementation effort, and then apply this to the unique types of values and beliefs that are associated with RTI in particular.

Culture and Systems Change

The more involvement and social cohesion within an implementation effort such as RTI the more likely real and sustainable change will occur (Walker et al. 1996). The change process itself occurs at the individual level and is considered a personal experience (Englert et al. 1993; McCarthy 1982). Hall and Hord (1987) describe stages of concerns that many individuals go through when participating in systems change efforts. These stages begin with general awareness and information gathering, with a focus, first, on how change efforts will impact each person on an individual level. This personal view of how a new practice will create change tends to occur before an educator is able to discuss and process any concerns related to the management of a new evidence-based practice and how these practices will impact students. At the end of this change process, when an individual emerges with a different understanding of reality, cultural change that is directly related to the implementation effort is achieved.

Culture and RTI

There are commonly shared viewpoints associated with RTI for both academics and behavior that help to create a culture of prevention. Examples of values and beliefs that are considered essential within RTI include the belief that all children can learn and succeed socially (Batsche et al. 2005). One way in which this value results in systems change is when changes occur with the traditional pre-referral systems used to evaluate whether students may need special education services. Schools where RTI is implemented successfully reorganize pre-referral activities to ensure that: (a) screening for students at risk occurs early, (b) intervention is provided matched to needs, and (c) that school personnel are organized to problem-solve so that adjustments in core and intervention can be made in a timely fashion when data indicate a change is warranted. Moreover, success is measured in increases in academic success as well as, or instead of, decreases in referrals for special education identification evaluations.

In order to change the beliefs any one teacher may have about why they would refer a student for support, early implementers must first focus on sharing the important features of RTI in a more generic format. The idea that a continuum of evidence-based interventions can be an efficient way in which to implement interventions is another cornerstone in many definitions of RTI (Batsche et al. 2005). Layered prevention efforts are intended to save time and increase efficiency within a school environment by first implementing interventions for all students, and early identification of students who are academically or behaviorally at risk for failure (Freeman et al. 2006). The number of more intensive interventions may decrease within this prevention model and more effortful and complex interventions will be fewer in number. By implementing more general interventions first, the school provides a more consistent and stable setting for students in need of individualized interventions (VanDerHeyden et al. 2007). Prevention, problem-solving, and early planning are valued behaviors that decrease the need for more highly effortful interventions.

District and building administrators who consider consensus building and group decision-making as essential to early RTI efforts hold an important value often articulated in both systems change and RTI research efforts (Adelman and Taylor 2003, 2007; Batsche et al. 2005). A school culture that includes a respect for and emphasis on organizational learning will have systems in place so that district and building staff members can easily review data regularly. Teams emphasizing organizational learning also place value on creating a working environment where individuals feel free to discuss failures (Fullan 1991). It is believed that these types of value-driven systems may result in implementers who are more successful in early RTI implementation. The goal of early implementation is to establish, over time, sophisticated interaction patterns among all school members where demanding processes produce social cohesion and progressive ideas assist in systems change.

Positive Behavior Interventions and Supports as a Model

A model that can be helpful for district and building teams new to RTI implementation was first reported in the literature on RTI for behavior by the National Technical Assistance Center on Positive Behavioral Interventions and Supports or PBIS (Sugai et al. 2000). The PBIS blueprint for school-wide positive behavior support describes how districts and schools organize problem-solving by focusing on data, systems, and practices (see Fig. 1). The PBIS model in Fig. 1 first establishes that the goal is for every child to achieve academic and social success.

To ensure academic and social success, districts and schools must use existing data to inform decision-making. New software systems that provide easy and efficient displays of data are important tools for new teams. Providing efficient access to data will increase a teams' focus on problem-solving and analysis, thereby helping to create a proactive learning organization. School teams can use data (patterns related to office referrals, detention, suspensions, expulsions, etc.) to decide what social expectations to focus on, where interventions should be implemented, and what modifications are necessary to improve implementation. Once data-based decision-making is in place, school teams can begin to make more objective and informed decisions about the school's academic and behavioral practices.

In addition to supporting decision-making, systems are needed to ensure student and staff behavior change is reinforced and supported over time. Part of these supports includes creating meeting infrastructures for ensuring the success of RTI efforts. Integrated problem-solving systems are designed to ensure that academic and behavioral data are communicated and shared to inform effective practices. Student outcomes can be changed by introducing new practices implemented by the adults within the building. Strategies for ongoing professional development, reward and recognition systems for staff, and public celebrations could be used to encourage sustainable implementation of RTI efforts (Herman et al. 2008). The next section explores the important data, systems, and practices necessary for the effective launch of RTI in districts and schools. New implementers with a guideline for changing cultural practice through consensus-

Fig. 1 A systems change model for launching response to intervention emphasizing data, systems, and practices

based decision-making processes that emphasize the use of data to inform academic and behavioral practices are shown in Fig. 1.

District and school leaders support decision-making, adopt and support practices, and design staff support mechanisms through intentional planning. The importance of these three areas of focus cannot be emphasized enough. Implementation practices and systems involved with RTI changes behaviors of all individuals within the organizations. As you can see from the systems change and cultural information mentioned earlier, the values, beliefs, and perspectives of all school personnel and other important stakeholders must be considered throughout the change process. Fixsen et al. (2005) conducted a meta-analysis of evidence-based practices implemented across a wide range of projects and programs, in part, to identify and summarize key issues related to changing the cultural beliefs and practices within schools and other organizations. While not narrowly intended for an educational audience, their work focuses teams on issues specific to implementation that result in improved outcomes (Fixsen et al. 2005). The resulting synthesis defined and outlined stages of the implementation process that are used here to assist new implementation sites as they design their tiered system. According to Fixsen et al. (2005), the process of implementation appears to follow predictable and distinct stages of implementation, with evidence-based practices such as RTI. The stages of implementation applicable to this chapter include what Fixsen and colleagues call exploration and adoption, program installation, and initial implementation (see Table 1 below). The table depicts the stages of implementation, the purpose associated with each stage, and the result of each stage that would transition organizations through the implementation process.

Innovation and Sustainability as Essential Issues During Early Stages of Implementation

The stages of implementation that set up sustainability to be addressed by new implementation sites include exploration and adoption, program installation, and initial implementation. These stages remain the focus of the chapter and are described below.

Table 1 Description of the stages of implementation

Stages of the implementation process	Purpose	Result
Exploration and adoption	Assess the match between identified needs and resources in order to make a decision about whether to proceed	Decision is made whether to proceed with implementation
Program installation	Carefully attend to infrastructure decisions prior to implementation	Ensure details have been carefully considered and planned for prior to implementation
Initial implementation	Individual and organizational implementation of new programs, routine, and/or practices with a focus on learning about implementation	Launching new practices or programs with feedback loop to direct needed support
Full operation	Expand implementation efforts and adjust based on lessons learned during initial implementation	Implementation of protocols and practices are executed as designed with a high degree of integrity
Innovation	Expansion and refinement of the protocol based on lessons learned after implementation with integrity	Modifications made to improve or streamline programs, practices, or protocol after experience and high levels of implementation integrity
Sustainability	Demonstration that implementation efforts have become so engrained as practices that they are viewed as standing practices	Evidence of long-term implementation and effectiveness despite a variety of changes that occur naturally over time

Exploration and Adoption

One of the central tasks during exploration and adoption is to recognize and honor the leadership and planning that is required prior to implementation. Recognizably, district and building leaders want to see progress by creating movement towards implementation so that the change process becomes visible. A series of blueprints were created through the National Association of State Directors of Special Education to be used in state, district, and building teams (Kurns and Tilly 2008; Elliott and Morrison 2008). These blueprints were designed to lead teams through the broad decisions that would lead to implementation. In the blueprints, key actions and more detailed steps that districts and buildings would address during exploration and adoption demonstrate to teams both the scope of the work that will be involved, and serve as a basis for subsequent needs assessments. The following table provides important critical actions that teams would be addressing during exploration and adoption along with considerations that have been found to influence both launch and sustainability (see Table 2).

As depicted in Table 2, the exploration and adoption stage centers around creating consensus at multiple levels within a system, articulating a transparent process used to evaluate entry status of systems, data, and practices in order to plan next steps, and utilizing a decision-making process that acknowledges the multifaceted issues related to successful implementation.

Creating consensus is an embedded expectation for all the actions outlined in Table 2. Consensus is necessary for a culture ready to embrace creating a durable tiered system. To implement systems change within a district or building, all stakeholders must be involved in a meaningful way (Curtis and Stollar 2002). Some of the key actions in Table 2 involve engaging stakeholders in meaningful conversations regarding the underlying core principles of RTI, establishing a common rationale for adoption, and determining nonnegotiables for implementation. At the helm of the actions and decisions, administrator buy-in is essential because active leadership is needed

Table 2 Example district and building actions and considerations during exploration and adoption

Action	Considerations
Identify representatives for leadership team	Ensure representation across district/building departments (e.g., academic specialists, behavioral expertise, special education, English language learner, Title 1), union leaders, information services, etc.
Examine the research regarding practices involved with a tiered system of support	Review the technical adequacy of assessment practices related to screening, progress monitoring, diagnostic, and outcome measures
	Attend to research from credible sources (e.g., peer-reviewed journals, IES Practice Guides, What Works Clearinghouse)
Discuss underlying core principles of RTI	Create opportunities for candid discussions about these principles in relation to moving forward (e.g., the core principle that can be effectively taught to all children is one that cannot be assumed among committee members or faculties)
Examine current practices, priorities, and initiatives in relation to those involved with RTI	Attend to RTI practices that are aligned, a natural fit with existing practices, and those that would create tension
Establish rationale for adoption of an integrated tiered system of support	Attend to connection of tiered practices to federal and state laws (e.g., ESEA, IDEA, School Improvement Plans)
	Make connection with district data, goals, priorities, initiatives
	Create a vision of potential impact for students, families, educators, building/district
Determine non-negotiables	Based on decisions derived from consensus in this phase, articulate known non-negotiables. This will be revisited during initial installation

IES Institute of Education Sciences, *ESEA* Elementary and Secondary Education Act, *IDEA* Individuals with Disability Education Act during all stages of RTI (Ervin et al. 2006; Fullan 2005; Horner et al. 2005). In fact, administrator buy-in and leadership appear across many types of system change efforts including RTI (Fixsen et al. 2005; McIntosh et al. 2014; Sindelar et al. 2006).

Implementation readiness by key stakeholders is more likely when there are strategies for ensuring commitment by each person. The commitment means that individuals agree that RTI is a priority within a district and school, and that everyone is prepared in advance to view the undertaking of RTI as a long-term, multiyear investment (Horner et al. 2005). For example, the Michigan's Integrated Behavior and Learning Support Initiative requires that districts demonstrate support moving into the state training system through a signed agreement between the superintendent and statewide project directors (Goodman and Miller 2012). In turn, some schools follow this model by securing a signed agreement by all faculties indicating consensus for moving forward with planning, training, and implementation. For change to occur, there must be leaders across all levels within a system (Fullan 1991). Administrators, teachers, counselors, bus drivers, custodians, and students are important leaders within RTI efforts. Sustainability is strengthened when collective consensus is built, nurtured, and promoted by preventing change efforts to be perceived or tied narrowly to a single leader or department within the organization.

In many districts and buildings, preparation during exploration and adoption focuses on nonnegotiable and negotiable requirements for the implementation of RTI (Fixsen et al. 2005). At the district level, decisions regarding core curriculum and time allocations may be nonnegotiable. However, each building may be able to determine their individual master schedule. Similarly, the district leadership team may determine specific reasons why they are advancing the creation of an integrated academic and behavioral tiered system, but each building needs to discuss and add their own issues and anticipated outcomes that they seek to improve for their building. This process of achieving initial consensus as to why an integrated RTI process is being adopted and what it will involve related to assessment, curriculum, and instruction needs to be achieved at the district and building level. The process will be formatively evaluated and revisited during initial implementation and subsequent phases.

During exploration and adoption, teams need to use a transparent and open process for evaluating current status of systems, data, and practices in order to plan next steps. While individuals may have a process for evaluation, district or building teams will be well served to discuss and develop consensus around the tools and process that will lead to collective decision-making. For example, when examining assessment practices, teams need to be clear about what it means to select valid and reliable tools for screening decisions. As is emphasized in Table 2, teams need to review the technical adequacy of potential measures. Districts and buildings cannot assume that a purchased assessment tool meets acceptable levels of reliability and validity (Kameenui et al. 2006). Further, from a behavioral perspective, teams need to understand state reporting requirements related to behavioral infractions and determine the impact on building-level determination of majors and minors when developing their school-wide behavioral system.

A case in point that has been seen in several districts is the adoption of web-based reporting systems for monitoring and tracking data that is needed for decision-making in a tiered model. Without proper planning, teams found themselves off the ground with implementation and then having to double enter academic and behavioral data because of state reporting that was not discussed, and an academic system that was being built by a department that was not represented on the district team. The culture of districts is that collaboration and communication are intended, but formal structures are necessary to ensure the voice of all stakeholders is present.

One of the most powerful areas for teams to explore during exploration and adoption is that of evidence-based practices (Fixsen et al. 2005). High-yielding instructional practices provide a pivotal point of impact related to positive, active engagement, and increases in learning. When examining the research to determine high-yielding practices, Stanovich and Stanovich (2003) offered the following guidelines for being wise consumers of research: (a) determining whether the research has been published in a peer-reviewed journal or approved by a panel of independent experts, (b) seeking evidence that the practices or programs have been replicated by researchers, and (c) determining whether there is consensus in the research community that sufficient evidence exists to support the practices. These criteria assist districts and teams with a robust way to navigate educational research. Following guidelines such as these helps organizations to make the strongest decisions regarding evidence-based practices. As noted in the work of Hattie (2009), the problem with research in education is that almost all effect sizes from the research in education are positive. Rather, what is important when building a strong and durable system is that the practices selected are those that produce real and noticeable gains.

When districts and buildings are making decisions about evidence-based practices, decisions made revolve around instructional materials, design, and delivery. Evaluating instructional materials could include evaluating particular instructional programs that are being considered for adoption (e.g., a reading series, social skills curriculum), or critiquing instructional materials that might be selected to teach a particular skill (e.g., evaluating the appropriateness of text selected to support a reading technique). Teams need to have clear expectations for their review, in terms of both critical features they are evaluating and a scale for evaluation. For example, Coleman and Pimentel (2012) provide criteria that can be used to evaluate the alignment of language arts materials with the Common Core State Standards. These criteria provide a focus for teams as they review materials such as evaluating the degree to which the materials adequately address academic vocabulary that is encountered in complex texts, ensure that all students have extensive opportunities to encounter and engage with grade-level text, and the degree to which they build in scaffolds that enable all students to experience rather than avoid the complexity of the text.

When evaluating the research related to instructional delivery, teams are recommended to pay careful attention to high-yielding practices that the teacher demonstrates in order to teach and reinforce learning objectives. Examples of teacher delivery behaviors would include the extent to which learning objectives are taught with strong and explicit modeling, guided practice, and judicious review over time. Simmons and Kame'enui (2003) created an evaluation tool used to review core materials that evaluates instructional practices within the design of the materials. Tools that guide teams through a careful analysis of materials and keys them into important teaching practices positions their decisions to yield stronger student outcomes.

Teams will gain strong momentum and mileage if they begin implementation efforts by emphasizing the need to maximize instructional time and student engagement in an integrated fashion including the use of both academics and behavioral interventions. Research has demonstrated that strong first instruction, by itself, can reduce problem behavior (Filter and Horner 2009; Preciado et al. 2009). Using the work of Hattie (2009), teams can begin to identify and magnify the connection between teaching behaviors that have been demonstrated to be worth focusing on (e.g., big ideas in reading, clarity in expectations), those that serve to engage students (e.g., reciprocal teaching, cooperative vs. individualistic), and those that serve to provide feedback to the teacher about student mastery and understanding (e.g., formative evaluation). School teams can use the research base as a guide to make defensible decisions and promote purposeful practices.

While different decisions or areas explored may have independent tools that teams or committees utilize during evaluation (e.g., consumer's guide to evaluate an intervention program, evaluation matrix for technical adequacy of assessment instruments), a district or building team will want to ensure that they have an overall decision-making process that acknowledges the multifaceted issues related for successful RTI implementation. A tool developed through the National Implementation Research Network (Blase and Fixen 2013), provides six broad factors to consider during exploration and adoption. These six factors include: (a) needs of the students, (b) fit within the current district or building initiatives or priorities, (c) resource availability related to support, (d) evidence base to support the approach, (e) readiness for replication, and (f) capacity to implement.

When used by a district or building-level team, the discussions around these six areas not only assist with a thorough and well-rounded discussion of areas that impact overall implementation but also will serve teams when information is shared with stakeholders regarding what is being proposed, how decisions were made, and how the selected approach compared to alternatives under consideration. These factors are particularly important when the culture of trust has been compromised or is vulnerable. In addition, the contextual issues discussed become very important as teams move to detailed planning during initial installation.

Program Installation

Once a district or building has secured commitment and consensus to advance the planning into program installation, work commences related to building the proper infrastructure so initial implementation has a high probability of being launched successfully. If the team examined and discussed the areas represented in Table 2, then teams are well positioned to address infrastructure issues. As is illustrated in Table 3, formidable work is conducted prior to implementation.

One of the major decisions that impacts an RTI implementation effort is the way in which schools are exposed to all tiers of intervention and how academic and behavioral interventions are implemented. In some districts, academic and behavioral systems are implemented separately with one multi-tiered approach beginning first within a school (e.g., school-wide positive behavior and intervention supports) followed by training and supports for academic systems. Additionally, the scale of implementation may vary. It may make sense in one district to create a slow expansion plan, thereby providing time for demonstration schools to thrive. In other situations, a number of trainers may already be available leading to a faster schedule of implementation across the district. Irrespective of the approach and scale of implementation plan, leaders need to understand and build the infrastructure for an integrated system from the onset. This may mean ensuring that district or building experts for academic and behavioral systems plan together to make natural connections and determine common tools that will be created for teams.

While all these areas are important to address, the authors chose to elaborate on three areas that impact the durability and sustainability of RTI and impact the cultural response to change including integrating RTI practices into existing required or priority processes (e.g., school improvement plans), making a defensible decision-making protocol, and developing an integrated professional and support plan.

As depicted in Table 3, one of the tasks during program installation is to take the work done during exploration and adoption related to connections made between RTI tiered practices and other requirements, initiatives, or long-standing routines. In order to avoid the proverbial "layering" approach, teams should review various reporting materials (e.g., school improvement form, referral for assistance, office referral form), district or building literature (e.g., elementary handbook, student code of conduct), and district or building-wide training materials. This task, in and of itself, is considered a professional development task for those involved. They are able to think systemically about the impact and connection with the practices directly involved in a tiered system. A message this assists with for those outside of the leadership team is that the practices and systems being promoted are not envisioned to be a short-lived initiative. The result, too, reduces duplicative work and serves to increase productivity. For example, districts or buildings likely have some type of handbook that is provided, or is accessible, to school patrons. The leadership team needs to ensure that the decisions made regarding curriculum, instruction, assessment, and procedures for accessing assistance are accurately reflected

Table 3 Example district and building actions and considerations during program installation

Action	Considerations
Provide information about RTI to stakeholders	Identify internal and external partners who have a vested interest in RTI. While it is recognized that information will be shared at different times, make the identification of stakeholders broad (e.g., include agencies who serve district/building students like county mental health offices, family services)
	For sites that choose to begin planning related to a particular domain (e.g., academics) or in a particular content area (e.g., reading), present the vision of an integrated model from the onset
Integrate RTI practices with school improvement plans and other district initiatives	Demonstrate integration in all written materials
Develop a communication plan	Plan should be designed as feedback-looped system between district and building teams, or between staff and building leadership team
Identify funding sources that may be used to support RTI	Identify funding sources, parameters for use, and fiscal start/stop and carryover target dates
Conduct a gap analysis related to	Ensure a common language and understanding of the area and criterion for analysis (e.g., strong evidence base, sufficient technical adequacy for decision being made)
Universal instruction/expectations	
Assessment systems	
Tiered supports	
Teaming and problem-solving processes	
Determine decisions related to	Nonnegotiables should be revisited in light of the more specific decisions made in this phase
Universal instruction/expectations	
Assessment systems	
Tiered supports	
Teaming processes	
Special education evaluations	
Develop an evaluation process related to:	Evaluation process should be formative and include tools that inform district or building decision-making. They should not rely solely on self-assessment tools
Implementation	Calendars or electronic prompts are helpful for teams to ensure data collection
Impact on student outcomes	
Effectiveness of programs/practices	
Develop an integrated professional development and support plan related to	A full continuum of support should be articulated from basic awareness information to on-site coaching
Universal instruction/expectations	Longitudinal plans should be crafted thinking through how training and information will be provided to new staff members
Assessment systems	
Tiered supports	
Teaming and problem-solving	

RTI response to intervention

in this document. Likewise, reviewing documents like school improvement plans will be important for teams to understand how the information that will be gained from their RTI tiered system will impact the goals, activities, data, and staff development sections of the plan. When developing an integrated system, teams should consider adding a goal related to their work on student social competencies. In the state of Kansas, for example, school improvement plans need only address goals related to reading and mathematics. In districts with strong integrated models of RTI, a third goal around social competencies is added, as well as being integrated into the academic goals as increasing academically engaged time.

Developing the decision-making process related to universal instruction and expectations, assessment systems, tiered supports, and interface/impact on the special education evaluation process is no light undertaking. Of all these decisions, the decisions tied to how the assessment data are used within a tiered system is one that can hinder sustainability the most. Some of the assessment data decision rules that teams need to think through at the individual student level include: (a) what criterion will be used with the screening data that would indicate a student may need additional assistance, (b) how do one set goals for those receiving assistance, (c) how will one know if the core and intervention are sufficient, and (d) how do one know when to reduce or increase the supports offered? Different assessment systems often provide guidance for teams that would allow them to craft a defensible protocol. The issue not often tended to during this phase is how the decision rules foster, or promote, the important underlying question the data were designed to answer. When decision rules are developed and trained, the important message to be heard and understood is that the rule represents an indication of concern, progress, or indication that the team should consider an action. For example, teams need to understand that the data rule associated with interpreting screening data is that students performing at or below this level may need additional support. When decision rules do not emphasize this, the unintended outcome is that the decisions become a rote application of rule, rather than an important reflection and decision about a student. The important message for teams developing and training rules is to ensure that the message of intent needs to be what is heard, understood, and reinforced. All decision rules should be applied in the context of understanding the intent of the underlying rule. When this is not the case, districts often end up having rules replacing rules with no change in adult behavior, just a change in the rule followed.

A daunting reality is that the changes involved in creating a tiered system require a substantial amount of professional development. While professional development is the last action covered in Table 3, it represents very careful thinking and planning. Professional development that is sufficient for school personnel to be comfortable, confident, and able to display appropriate demonstration of the practices will involve providing a full continuum of supports. The seminal work of Joyce and Showers (1980) remains a current-day reminder that must create in-service skill development opportunities that are job embedded and provide coaching and collegial support. As Fullan (1991) cautioned, everyone must be careful not to over-assume the capacity of teachers to move actively into implementation without providing an adequate amount of support. With RTI-tiered practices often involving many new skills, and revisiting and sharpening many existing skills, the notion of going slow to go fast is an important factor. In order to get a handle on this, teams should follow a process as listed:

1. Develop a scope and sequence of trainings that will be involved with implementation.
2. Identify all district personnel and determine who “needs what and when.”
3. Determine existing professional development opportunities within the calendar year.
4. Based on existing professional-development opportunities, determine reality of what will be able to be covered on an annual basis.
5. Determine sufficiency of the plan—can more days be carved out in order to accomplish more?
6. For each in-service training provided, determine plan for building or extending local capacity and support plan for coaching. Coaching should be considered along a continuum that would range, for example, from phone consultation all the way to side-by-side coaching.

When professional development plans are launched, districts and buildings begin to see the fruits of their labor begin to take shape.

Initial Implementation

A key function of any professional-development system should be to unite staff within their school as they work together to implement positive change (Joyce and Showers 2002). As already mentioned, part of creating a “culture of change”

Table 4 Example district and building actions and considerations during initial implementation

Action	Considerations
Schedule regular meetings for the leadership team to: formatively evaluate implementation	Be attentive to opportunities to celebrate
	Provide time in agenda to discuss intended and unintended outcomes
	Revise action plans accordingly
	Identify and access internal or external support
Review progress	Maintain list of future areas of focus (e.g., application of practices for students with significant cognitive disabilities, developing overview trainings for substitute pools)
Manage resources	
Problem-solve and adjust implementation plans	
Implement the communication plan as designed	Post district responses to questions on internal server to reduce misunderstandings and provide clarification
	Be attentive to groups that may not be represented on leadership team
	Add reflection agenda item to all standing district committees or councils
Provide on-site support and mentoring	Provide internal and/or external coaching during initial launching of key practices (e.g., observing or facilitating data reviews, observing or coaching during classroom or intervention instruction) so that early feedback can be provided
	Utilize internal and external coaching meetings to identify areas in need of coaching/support
Evaluate implementation fidelity	Utilize self-evaluation tools, direct observation, and permanent product review in order to determine if practices are being implemented as intended
	It is important that resulting data are used in a supportive manner

involves fostering collaborative teams within schools where learning, reflective processes, and collegial interactions are encouraged over a long-term basis (Glatthorn 1990; Hollingsworth 1992; Lieberman and Miller 2002). During initial implementation, it will be imperative that this culture is promoted early, reinforced often, and adequately supported around key actions outlined in Table 4. Teams are encouraged to attend to the initial implementation actions with a conscious emphasis placed on the degree to which the supports provided model the preventive and intervening aspects of RTI practices by those providing internal or external supports.

By promoting implementation of RTI as a collective experience with opportunities for sharing and reflection, individuals can depend and trust on support from each other and the district through the change process. An emphasis is added here on the change *process*. Once initial implementation is launched, leaders are reminded on a frequent basis that change is a process, and often a difficult one. The authors have had several experiences that bring teams through all the hard work of the previous stages, launching implementation plans, create initial momentum and excitement, only to be short lived and snapped back to the status quo. This occurs, in part, due to inattention to the elements of change that take careful consideration, patience, and persistence. It is the recommendation that teams meet often and review frequently their celebrations, address what is surfacing as concerns, and building the adequacy of supports that are needed for practices to be implemented.

The team approach provides an important communication hub for the implementation process and become a source of ongoing learning. When a district or school has both formal and informal methods for communication about implementation efforts that converts tacit knowledge to explicit knowledge, teams are able to tap into values, meanings, day-to-day skills, and experiences of everyone in the school making this information available for problem-solving (Fullan 1999). One of the considerations highlighted in Table 4 is that of responding promptly to questions or concerns. In the absence of clarification, misinformation can quickly be spread as fact and result in undesirable student and adult outcomes.

A barrier noted by many engaging in systems change and RTI implementation relates to failure to establish ongoing coaching, training, and supports for school personnel (Fullan 2005; Joyce and Showers 2002). Comprehensive, longitudinal professional development systems are better able to address the developmental pace of learning (Colvin et al. 1993) and provide opportunities for school staff to engage in collaborative dialogue, feedback, and reflection about their practices (Guskey 1995; Ellmore and McLaughlin 1988). When these opportunities are not provided to teachers, a common response is to fall back to earlier learning histories: "When faced with too many challenges and discord between new knowledge and existing knowledge, it is sometimes easier to revert back to known and familiar patterns (p. 413, Klingner et al. 2003)**."**

Conclusion

The purpose of this chapter was to discuss key challenges for new teams using Fixsen's first stages of implementation as described in Table 1. The way in which new implementers of RTI address these first stages of implementation will have an enormous impact on a team's ability to successfully move towards later stages of implementation including the stages Fixsen and colleagues refer to as full operation, innovation, and sustainability. In fact, when problem-solving systems have been successfully established, the foundations for successful implementation, innovation, and sustainable RTI naturally evolve. For instance, consider a school team that has included the cultural value of academic and behavioral RTI integration during the very early stages of implementation. The school team who began by reviewing the data for both academic and behavioral interventions as separate practices can evolve their systems to know when and how to examine the possible integrated function. The process becomes more efficient in the manner in which problem-solving occurs at the same time that it becomes more sophisticated. Basically, less time is spent interpreting the data and more time is spent analyzing and responding to the data. When teams reach a point where a function-based orientation and problem-solving process is in place, a new culture has been created.

As mentioned earlier, the practices and structures involved with RTI go beyond the implementation of a new strategy, data system, or referral process. The practices with RTI touch all aspects related to schooling: evidenced-based curricular and instructional practices, use of the problem-solving process driven by data, and teaming at different levels. Clearly, the meta-analysis conducted by Fixen et al. (2005) has provided important work regarding the many facets of implementation to which the efforts need to be directed. We also remain encouraged by the work examining the sustainability of school-wide positive behavior support and hope that it will be expanded to include academic structures as well (McIntosh et al. 2014; McIntosh et al. 2013).

It has been observed that buildings and schools which focus on the important values and beliefs associated with RTI early on in implementation can set the stage for an exciting and rewarding experience across each of the stages of implementation described by Fixsen and his colleagues. It is believed that early attention to a consensus-building approach using the data, systems, and practices model described in Fig. 1 helps new teams navigate early challenges associated with implementation. The successful implementation of RTI will depend greatly on the extent to which districts and schools are able to build the social cohesion and collegial problem-solving systems that are necessary, while at the same time supporting future research towards factors related to sustainability which help navigate this complex change process.

References

Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation, 14,* 1–26.

Adelman, H. S., & Taylor, L. (2007). Systemic change and school improvement. *Journal of Educational and Psychological Consultation, 17,* 55–77.

Batsche, G. M., Elliott, J., Graden, J., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria: National Association of State Directors of Special Education.

Blase, K., & Fixen, D. (2013). *The hexagon tool*: *Exploring context*. Chapel Hill: NC. National Research Implementation Network.

Burns, M. K. (2007). RTI WILL fail, unless…. *Communiqué, 35*(5), 38–40.

Coleman, D., & Pimentel, S. (2012). *Revised publishers' criteria for the common core state standards in english language arts and literacy, Grades K-2*. http://www.corestandards.org/assets/Publishers_Criteria_for_K-2.pdf. Accessed 13 May 2014.

Colvin, G., Kame'enui, E., & Sugai, G. (1993). Reconceptualizing behavior management and school-wide discipline in general education. *Education and Treatment of Children, 16,* 361–381.

Culture. [Def. 5]. (n.d.). *Merriam-Webster Online*. In Merriam-Webster. http://www.merriam-webster.com. Accessed 1 Nov 2012.

Curtis, M. J., & Stollar, S. A. (2002). Best practices in system-level change. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology: IV* (pp. 223–234). Bethesda: National Association of School Psychologists.

Elliot, J., & Morrison, D. (2008). *Response to intervention blueprints*: *District level edition*. Alexandria: National Association for State Directors of Special Education.

Elmore, R. F., & McLaughlin, M. W. (1988). *Steady work: Policy, practice and the reform of American Education*. Santa Monica: RAND.

Englert, C. S., Tarrant, K. L., & Rozendal, M. S. (1993). Educational innovations: Achieving long-lasting curricular change through extended collaboration and reflective practice. *Education and Treatment of Children, 16,* 441–473.

Ervin, R. A., Schaughency, E., Goodman, S., McGinchey, M. T., & Matthews, A. (2006). Merging research and practice agendas to address reading and behavior school-wide. *School Psychology Review, 35,* 198–223.

Filter, K. J., & Horner, R. H. (2009). Function-based academic interventions for problem behavior. *Education and Treatment of Children, 32,* 1–19.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, R. (2005). *Implementation research*: *A synthesis of the literature*. Tampa: National Implementation Research Network.

Freeman, R., Eber, L., Anderson, C., Irvin, L., Bounds, M., Dunlap, G., & Horner, R. H. (2006). Building inclusive school cultures using school-wide PBS: Designing effective individual support systems for students with significant disabilities. *Research and Practice for Persons with Severe Disabilities, 31*(1), 4–17.

Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Fullan, M. G. (1999). *Change forces: The sequel*. Philadelphia: Falmer Press.

Fullan, M. (2003). *Change forces with a vengeance*. London: Routledge Falmer.

Fullan, M. (2005). *Leadership and sustainability*: *System thinkers in action*. Thousand Oaks: Corwin Press.

Glatthorn, A. (1990). *Supervisory leadership*. New York: Harper Collins.

Goodman, S., & Miller, D. (2012, October). *A district model for integrated RtI systems*. Presentation at the 2012 National PBIS Leadership Forum, Rosemont, IL.

Guskey, T. R. (1995). Professional development in education: In search of the optimal mix. In T. Guskey & M. Huberman (Eds.), *Professional development in education*: *New paradigms and practices* (pp. 114–131). New York: Teachers College Press.

Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process*. New York: State University of New York Press.

Hattie, J. (2009). *Visible learning*: *A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hawken, L. S., Vincent, C. G., & Schumann, J. (2008). Response to intervention for social behavior: Challenges and opportunities. *Journal of Emotional and Behavioral Disorders, 16,* 213–225.

Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., & Darwin, M. (2008). *Turning around chronically low-performing schools*: *A practice guide* (NCEE #2008-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hollingsworth, S. (1992). Learning to teach through collaborative conversation: A feminist approach. *American Educational Research Journal, 29,* 373–404.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Joyce, B., & Showers, B. (1980). Improving inservice training: The messages of research. *Educational Leadership, 37,* 379–385.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed., p. 78). Alexandria: ASCD.

Kameenui, E., Fuchs, L., Francis, D., Good, R. III, O'Connor, R., Simmons, D., Tindal, G., & Torgesen, J. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher, 35,* 3–11.

Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children, 69,* 411–429.

Kurns, S., & Tilly, W. D. (2008). *Response to intervention blueprints*: *School building level edition*. Alexandria: National Association for State Directors of Special Education.

Lieberman, A., & Miller, L. (2002). Transforming professional development: Understanding and organizing learning communities. In W. D. Hawley & D. L. Rollie (Eds.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 74–85). Thousand Oaks: Corwin.

Lohrmann, S., Martin, S. D., & Patil, S. (2013). External and internal coaches' perspectives about overcoming barriers to universal interventions. *Journal of Positive Behavior Interventions, 15,* 26–38.

McCarthy, B. (1982). Improving staff development through CBAM and 4MAT. *Educational Leadership, 38,* 20–25.

McIntosh, K., Mercer, S. H., Hume, A. E., Frank, J. L., Turri, M. G., & Mathews, S. (2013). Factors related to sustained implementation of school-wide positive behaviour support. *Exceptional Children, 79,* 293–311.

McIntosh, K., Predy, L., Upreti, G., Hume, A. E., & Matthews, S. (2014). Perceptions of contextual features related to implementation and sustainability of school-side positive behavior support. *Journal of Positive Behavior Interventions, 16,* 29–41.

Preciado, J. A., Horner, R. H., Scott K., Baker, S. K. (2009). Using a function-based approach to decrease problem behaviors and increase academic engagement for Latino English language learners. *The Journal of Special Education, 42,* 227–240.

Simmons, D. C., & Kame'enui, K. J. (2003). *A consumer's guide to evaluating a core reading program Grades K-3*: *A critical elements analysis*. http://oregonreadingfirst.uoregon.edu/inst_curr_review.html. Accessed 13 May 2014.

Sindelar, P., Shearer, D., Yendol-Hoppey, D., & Liebert, T. (2006). The sustainability of inclusive school reform. *Exceptional Children, 72,* 317–331.

Stanovich, P. J., & Stanovich, K. E. (2003). *Using research and reason in education: How teachers can use scientifically based research to make curricular and instructional decisions*. Portsmouth: RMC Research Corporation.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R. III, Wickham, D., Ruef, M., & Wilcox, B. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2,* 131–143.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., et al. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional & Behavioral Disorders, 4,* 194–209.

Making Response to Intervention Stick: Sustaining Implementation Past Your Retirement

Kim Gibbons and W. Alan Coulter

Educators in the twenty-first century are now required to do something that has never been expected of them before: ensure high levels of learning for all students (DuFour 2004). However, the National Assessment of Educational Progress (NAEP) scores indicate a need to accelerate student growth in both reading and mathematics. The average fourth-grade reading score in 2011 remained unchanged from 2009, but was 4 points higher than in 1992. The average eighth-grade reading score in 2011 was 1 point higher than in 2009, and 5 points higher than in 1992. At both grades 4 and 8, the percentages of students performing at or above the basic level, did not change significantly from 2009 to 2011, but were higher in 2011 than in 1992 (Eric et al. 2010). The data are similar for mathematics, with both fourth- and eighth-graders scoring higher on the NAEP in 2011 than in previous assessment years. However, despite this positive news, a recent report on achievement growth within the USA and internationally indicates that only 6 % of US students are at the advanced level in mathematics, and only 32 % of 8th graders are proficient in mathematics (Hanushek et al. 2010). These statistics place the USA 32nd when ranked internationally.

K. Gibbons (✉)
St. Croix River Education District, Rush City, MN, USA
e-mail: kgibbons@scred.k12.mn.us

W. A. Coulter
TIERS Group, Human Development Center,
LSU Health Sciences Center—New Orleans,
New Orleans, LA, USA

The response-to-intervention (RTI) or multi-tiered systems of supports (MTSS) framework has quickly emerged as a methodology for improving outcomes for all students through high-quality instruction tailored to student needs within a data-based decision-making model. In fact, a recent national survey of K–12 administrators indicated that 61 % of respondents are either in full implementation or in the process of district-wide implementation of an RTI/MTSS framework, up from 24 % in 2007 (Spectrum K–12 Solutions 2010). Alexa Posny, Assistant Secretary for Special Education and Rehabilitative Services stated that, "RTI and EIS (early intervening services) are absolutely the future of education—not the future of special education, but of education" (p. 9, Project Forum 2007).

While it is promising that so many school districts around the country are beginning to implement the RTI/MTSS framework, there has been less written in the research literature on how to sustain implementation over the long term. The field of education is full of examples of school districts implementing an innovation with great enthusiasm at the beginning, but then later abandoning it (Fullan 2007; Levin et al. 2005, Payne 2008), and RTI/MTSS could be the latest in the long list of programs that were 'tried years ago.' Overcoming this challenge is critical because RTI/MTSS represents the best chance for substantive change since P.L. 94–142 was first passed (Burns and Gibbons 2012).

Recognition that initiatives often fail has led to studies building a body of knowledge on the

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_36

processes of implementation. There is a growing acknowledgement that installing a new program or initiative requires both a thorough knowledge of what is to be initiated and attention to the context in which the new program will be implemented. This body of knowledge is called *implementation science*, which is the scientific study of methods to promote the uptake of research findings into routine settings in clinical, community, and policy contexts (Fixsen et al. 2007). Similarly, the National Institute on Health (NIH) refers to implementation science as "the study of methods to promote the integration of research findings and evidence into policy and practice" (p. 1, NIH 2013).

The key to any successful educational reform effort, including RTI/MTSS and MTSS, is to ensure sustainability over time. The problem for educators has now become not how to initiate RTI/MTSS; rather, it is how to implement changes in process and performance in a manner that endures. It is known that issues will arise as implementation occurs on a national level within individual districts. While it is impossible to foresee every potential problem, one can predict themes that will occur if key components are not addressed or not implemented with fidelity. The purpose of this chapter is to discuss how educators and systems can institutionalize RTI/MTSS in ways that preserve the positive changes and instill resilience in resisting efforts to revert back to the old ways of doing things in schools.

Sustaining an RTI/MTSS Framework

As with other systemic improvement efforts, implementing an RTI/MTSS framework requires system change on many levels (Kratochwill 2007). Some researchers have suggested that school districts will have the best probability of implementation success if they implement the RTI/MTSS framework in phases over time, initially focus on enhanced learning rather than making entitlement decisions, focus on tiers 1 and 2 in addition to tier 3, operationally define RTI/MTSS, measure fidelity of implementation, and match interventions to the individual needs of students (Burns and Gibbons 2012; Burns 2007). It is suggested that in order to effectively manage and sustain the change process, districts need to attend to seven imperative components: (a) Leadership, (b) Vision and Culture, (c) Infrastructure, (d) Resources, (e) Implementation Plans, (f) Professional Development (knowledge, skills, and self-efficacy), and (g) Incentives. The

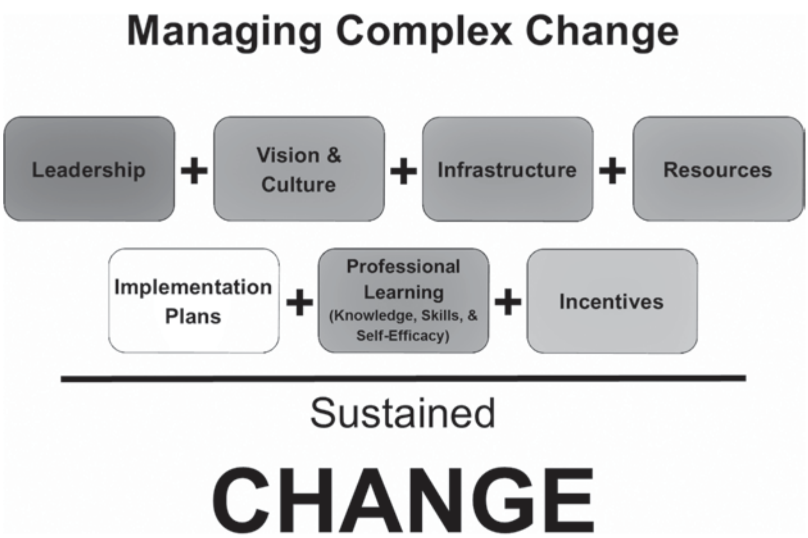


Fig. 1 Managing complex change

components are listed in Fig. 1. If any of these components are not addressed, the change process will be impacted (Knoster and George, 2002). The next section outlines each of these seven and discusses how each area impacts the change process and successful implementation.

Leadership

State education agencies and local districts set the context and tone for implementation of RTI/MTSS through their leadership in policies and support of professional development (Shelton 2011). Local school superintendents lead by building consensus and selecting key school staff (Hoerr 2005, Reeves 2005). In the past decade, research has supported the central role of leadership, particularly the instructional leaders (building principal or school leader), as a key factor in improving student outcomes. Of the priorities listed for school reform, improving school leadership ranks second to teacher quality (Seashore et al. 2010). Most school variables have a small effect of student learning, but an effective leader works to combine individual variables to reach a critical mass and a tipping point for implementation. As a result, the foundation of educational reform efforts are principals who move from the role of manager to the primary role of instructional leader using data (McNulty and Besser 2011) Principals need to become leaders of learning who can develop a team culture of using data and delivering effective instruction (Marzano et al. 2005).

The Wallace Foundation (2013) suggests five key responsibilities of instructional leaders. First, they must communicate a vision of high standards for all students. Second, they must create a hospitable educational environment that focuses on collaboration among staff to embrace high standards and results. Third, they must encourage leadership in others, so that the entire school community is results-driven. Fourth, the primary building focus must be on improving instruction. Finally, instructional leaders must have the ability to manage people, data, and processes to improve outcomes for their students. Meeting these responsibilities will require that instructional leaders are results oriented and data informed. They must be responsive to feedback and be effective communicators. Above all, they must work to build consensus and articulate a vision for the building that aligns with district priorities.

Vision and Culture

Countless books on leadership exist that highlight the importance of leaders who have vision (Honore' 2012; Jackson and McDermott 2012; Reeves 2005). President Kennedy dreamed of putting a man on the moon. Eleanor Roosevelt envisioned a world of equal opportunity for women and minorities. Great leaders think about the values and ideas that they are most passionate about and actively work to put their vision into action. They realize that putting vision into action is only possible through setting ambitious goals and then mobilizing their teams to meet these goals. Senge (1990) hypothesized that vision and clear statements of future benefits for the system and people it serves are essential to reform. Furthermore, conveying a clear vision and the intended benefits of reform efforts along with a clear plan of the supports that will be provided during the change process is thought to contribute towards readiness and acceptance (Fullan 1991).

Talking about vision as it relates to implementation of the RTkI/MTSS framework, district leaders need to be cognizant of how RTI/MTSS relates to the district's strategic plan and goals. Efforts need to be made to explicitly communicate how RTI/MTSS fits within other district initiatives (e.g., braiding initiatives). For example, many districts are implementing professional learning communities (PLCs), benchmark assessments for all students, progress monitoring for some students, positive behavior interventions and supports (PBIS), anti-bullying programs, standards-driven instruction, and pay-for-performance systems. A leader with good vision will explicitly show how each of these initiatives fit within an RTI/MTSS framework and are related to each other. Without a clear vision, staff members will be confused. A set of guiding ques-

Table 1 Guiding questions for establishing vision

Guiding questions to help establish vision
How is RTI related to the district's strategic plan or goals?
Is it explicitly communicated how RTI fits within other district initiatives?
Is there agreement on a common measurement system for screening and progress monitoring across the district that is reliable and valid?
Is there a practice of databased decision-making at all levels?
Are data used to evaluate district implementation and evaluate grade levels to target for support?
Other questions??

RTI response to intervention

tions to assist district leaders with setting a clear vision may be found in Table 1.

School culture is defined as "the traditions, beliefs, policies, and norms within a school that can be shaped, enhanced, and maintained through the school's principal and teacher-leaders" (Ward 2004, p. 1). Similarly, Deal and Peterson (1998) view school culture as "underground flow of feelings and folkways wending its way within schools in the form of vision and values, beliefs and assumptions, rituals and ceremonies, history and stories, and physical symbols (p. 28). School cultures reflect a community of practice where integration of function and mutual respect is paramount" (Sailor 2009, p. 188). To ensure enduring change, RTI/MTSS must accommodate school culture and adjust to the unique features of a district and each of its buildings. Sailor (2009) asserts that RTI/MTSS that is sustained modifies school culture by affecting how teachers and school leaders interact and implement their roles in delivering instruction.

Infrastructure

Building an infrastructure to support and sustain the RTI/MTSS framework is essential for successful implementation. Developing infrastructure involves identifying supports already in place for implementation of the framework, determining needs, designing or adopting a data management system, and creating policies and procedures to define implementation. It is suggested that districts attend to five key areas in developing infrastructure: (a) teams, (b) routines, (c) problem-solving process, (d) technology support, and (e) technical assistance.

Teams Teams are an essential foundation in building infrastructure. State agencies should operate a state leadership team that monitors state-level data and implementation of federal and state policies related to RTI/MTSS. Burns and Gibbons (2012) recommend that districts have four team support structures in place: (a) district leadership teams, (b) building leadership teams, (c) grade-level teams or professional learning communities, and (d) problem-solving teams. All these teams work toward a common mission of carrying out the district implementation plan and ultimately increasing student achievement. The district leadership team is responsible for district-level data analysis, consensus building, formulating a long-term implementation plan, explaining how other district initiatives fit within the RTI/MTSS framework, setting standards and methods for measuring fidelity of implementation, assisting buildings with a self-study of existing practices, and giving permission to abandon ineffective practices (planned abandonment).

Building leadership teams serve as the overarching management group for facilitating and evaluating implementation in a particular school. The building leadership team assists in completing a needs assessment to evaluate current assessment and intervention practices, identifies skill sets needed for training, and assists in building consensus and commitment. The primary role of this team is to evaluate school achievement and behavior data to identify needed changes in existing tools, training, and support, especially around fidelity of implementation.

Grade-level teams of teachers, the third type, meet regularly and collectively consider all students in a grade level as one group to be supported together (Allison et al. 2010; Sailor 2009). The grade-level team process is similar to the professional learning community (PLC) model (DuFour 2004, DuFour et al. 2010) or whole faculty study group (WFSG; Murphy 2006). However, it is different from previous approaches to professional development in which the teams selected a topic about which they wanted to learn more and engage in self-directed professional development (e.g., group book study). Instead, current PLCs, WFSGs, and grade-level teams embrace the notion that the purpose of school is learning and results, not teaching (DuFour 2004). Teams concentrate their energy on using data to find out (a) what students need to learn to achieve expected results, (b) how teachers will know when students have acquired intended knowledge and skills, and (c) how teachers will respond when students are discrepant from expectations (Ferriter et al. 2013). Zamuda (2004) terms the reoccurring process used by these teams as defining reality through data.

Strong grade-level teams are needed to increase student achievement through intensive data analysis, quality core instruction, and to develop well-designed standard treatment protocol group interventions. Problem-solving teams are needed to assist grade-level teams when their efforts do not result in student success at tier 1 or tier 2. The function of the problem-solving team (PST) is to design individualized and/or small group interventions using a problem-solving decision-making process to reduce referrals for special education services. However, it is critical that the PST be viewed as general education rather than a special education process.

Districts that lack team infrastructures will not sustain implementation of RTI/MTSS. If district and building-level teams do not exist, there will not be a structure to monitor and manage implementation and provide support to staff. If grade-level teams do not exist, problem-solving teams will be inundated with referrals for special education evaluation. Moreover, buildings without problem-solving teams will likely have high referrals to special education child-study teams.

Routines Schools have been described as loosely coupled systems (Weick 1976) that behave in routines to accomplish tasks and respond to authority. Organizational routines, an essential aspect of culture, involve "a repetitive, recognizable pattern of interdependent actions, involving multiple actors" (Feldman and Pentland 2003, p. 311). Routines or habits are built and maintained by schedules and school bureaucracy. To effectively change the way schools and teachers behave, routines must be changed. Routines that are critical to sustained RTI/MTSS include time allocated for team meetings, data collection on student knowledge and skills, collections of implementation data, data analysis by teachers, documented instructional planning, and a standardized problem-solving process. If school routines are not changed by the implementation of RTI/MTSS, then schools will perceive that what is being required is transient and, like other innovations, will soon pass so that the school can get back to "normal" (Sherer and Spillane 2011). For example, each building based problem-solving team should follow an established schedule and a specific process for responding to identified student concerns.

Problem-Solving Process A third area of needed infrastructure is the use of a standardized problem-solving decision-making process. It has often been stated that the reason most people are confused is that they do not know what question they are trying to answer. A problem-solving model guides implementation by framing essential questions and considering what data are needed to answer the question. Many problem-solving models exist; however, most models (a) identify a problem, (b) analyze why the problem is occurring, (c) develop a plan, (d) implement the plan and ensure that it is implemented with integrity, and (e) evaluate whether the plan worked (See Fig. 2).

A problem-solving model may be used at the systems level (e.g., building and district teams), group level (grade-level teams), and individual level (problem-solving teams). For example, the building-level team may use the problem-solving model to guide decision-making around whether core instruction is effective at various grade levels. Grade-level teams may use the model to develop standard treatment protocol interventions for students who need additional intervention to increase oral reading fluency. Problem-solving teams may use a problem-solving model to design an individ-

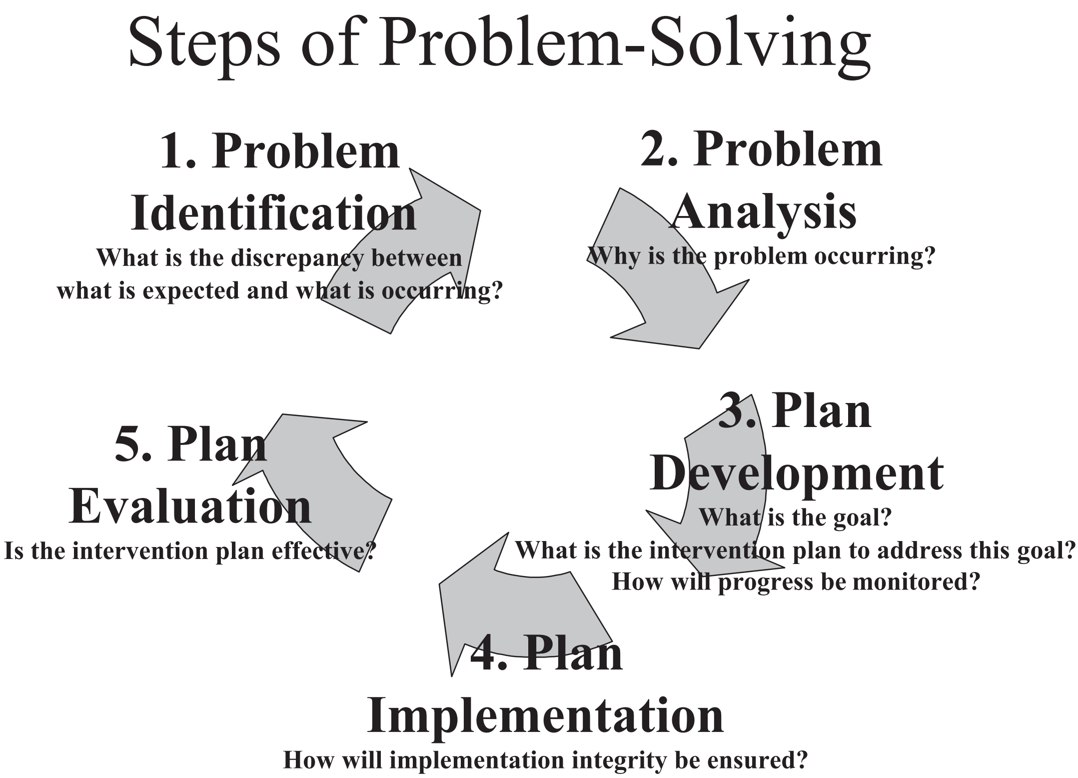


Fig. 2 Steps of problem-solving

ually tailored mathematical intervention specific to an individual student's needs. When the use of a problem-solving model for decision-making has been institutionalized, it allows all building members to speak a common language, facilitating efficient decision-making (Horner et al. 2010).

Technology Support A third area of necessary infrastructure is technology support. Technology support is, perhaps, the most underrated element of the infrastructure. Districts will need to develop or purchase a computerized database that stores, calculates, and displays student data for use by all teams within the district. In addition, some type of database should be in place to document tier 2 and 3 interventions and fidelity of implementation. Technology also plays an important role in communications and professional development. Thus, there should be support for (distance learning, blogs, wikis, etc.) tailored to the needs of groups of teachers.

Technical Assistance Finally, the last area of infrastructure is technical assistance. Technical assistance (TA) around an RTI/MTSS framework may take many forms, including TA delivered digitally. Most districts that have sustained RTI/MTSS implementation over many years have written policies, procedures, and guidance around the essential elements of RTI/MTSS implementation. However, written guidance is not enough. There needs to be a cadre of individuals who work to continually support teachers and other building and district staff in implementation efforts. It is not unusual to find RTI/MTSS coordinators, data coaches, instructional coaches, and PLC facilitators (to name a few) in most districts.

It is essential to have a position that coordinates these efforts and works to keep things integrated.

Resources

Without resources allocated to RTI/MTSS implementation, staff will experience a great deal of frustration. States can utilize federal and state discretionary funds to support implementation. At the district level, resources need to be aligned to match the needs and goals of the action plan. Districts and buildings will need to examine their needs and provide high-quality professional development in targeted areas. As discussed below, professional learning works best when it is job embedded and followed up with coaching and support (DuFour et al. 2010). Subsequently, districts may need to allocate resources to hire staff to help sustain implementation. It is suggested that the roles of RTI/MTSS coaches and staff should be fluid and change according to building and district needs over time. Resources also will need to be provided to purchase or develop a measurement system that will assist with screening, diagnostic, and progress monitoring decisions. Finally, resources may need to be allocated to work on standards alignment and purchase core, supplemental, and intensive curriculum programs.

Implementation Plans

Once the vision has been established and RTI/MTSS is integrated within the district's strategic plan, an implementation plan should be developed that describes the district and building implementation process over time. An implementation plan is directed at how to install or infuse RTI/MTSS in schools and includes a detailed listing of activities, costs, expected difficulties, and schedules that are required to achieve sustainable implementation (Fixsen et al. 2007). Without an implementation plan, districts and building will likely experience many false starts, and staff will be unable to see both the short-term and long-term plan for where the district and their school building is headed. Clearly developed implementation plans provide a context for explaining why activities are occurring and where they fit within the larger picture of school district operations. During any type of change process, lack of clarity in direction often creates stress for individuals whose practice is changing (Sarason 1996). Implementation plans that communicate the district's plan for implementing the RTI/MTSS framework may alleviate stress associated with change (Fixsen et al. 2007).

The Implementation plan should be a written plan that addresses the key areas of RTI/MTSS implementation. Examples of sample plans may be found at www.rtinetwork.org and www.nasde.org and in Appendix A. Buildings should give careful thought as to what level of implementation they are currently at, and plan for movement to the next stage. Recent research on implementation science suggests six stages of implementation: (a) Exploration, (b) Installation, (c) Initial Implementation, (d) Full Implementation, (e) Innovation, and (f) Sustainability (Fixsen et al. 2007). Table 2 describes each level of implementation.

Implementation of an RTI/MTSS framework is not an overnight process, and typically takes 3–5 years for full implementation (Fixsen et al. 2007). Thus, the written implementation plan should have short-term goals (current school year) and long-term goals (future school years) along with strategic activities, estimated completion dates, and a designated person who is accountable for each activity. It is important that districts develop reasonable action plans and try not to institute too many changes during one school year. There is no one blueprint for how expansive an action plan should be. Some buildings choose to limit the focus of implementation by grade level where consensus has been achieved (e.g., start with just kindergarten) or focus area (e.g., reading, mathematics, or behavior).

When developing the plan, the district or building team will need to define what is negotiable and what is nonnegotiable. Negotiables are practices that may be determined by individual buildings or grade levels based on needs. For example, a district may require that interventions

Table 2 Stages of implementation. (Fixsen et al. 2007)

Stage	Description
Exploration	Exploration involves a small team doing research to learn as much as they can about RTI in determining whether to implement such an approach
Installation	Installation begins when the decision to implement is made and continues until the first use of the innovation (this may include planning, assigning job responsibilities, determining how it will be organized, initial team building)
Initial Implementation	Initial implementation begins when as many teachers try to use new practices in their classroom
Full Implementation	Full implementation occurs when practices have been installed and most professionals are comfortable, with practices operating smoothly
Innovation	Innovation occurs after implementing the plan and involves refining and making improvements
Sustainability	Sustainability focuses on how to continue and sustain the practice over the long-term

RTI response to intervention

used with students are research based, but they may allow buildings to select research-based interventions that are aligned with their needs rather than mandate specific interventions district-wide. Nonnegotiables are areas the district or building hold constant and are not open to modification for buildings or classrooms. A common nonnegotiable is in the area of measurement (McNulty and Besser 2011). A district may determine the measurement system that will be used for screening and progress monitoring. All buildings would be required to use the same measurement system.

After the implementation plan is developed, there must be a strategy to communicate the plan with all district staff. When communicating the plan to staff, it is important to emphasize that the plan will be evaluated through its implementation, and that modifications may occur based on implementation issues and feedback from staff. The action plan should be viewed as a flexible plan that guides implementation rather than a rigid plan that cannot be modified.

Professional Learning (Knowledge, Skills, and Self-Efficacy)

One of the most significant changes when implementing an RTI/MTSS framework is the change pertaining to the professional practice of teachers and other support staff within the building. Successful implementation of the RTI/MTSS framework requires that educators need to be knowledgeable about collecting and interpreting student data, effective instructional strategies and differentiation of instruction, and matching instructional supports to student needs. Critically, educators must believe that the skills they have to implement RTI/MTSS will be effective with their students (Sanetti and Kratochwill 2009). Self-efficacy fuels implementation of RTI/MTSS and increases sustainability. Professional learning (the former term was professional development) must precede implementation in order to ensure that the staff has the skills necessary to be successful with this initiative and core beliefs that their efforts will be rewarded with improved student performance. Knowing the existing skills of the staff and what skills will be needed will inform professional development needs.

Perhaps the most critical skills educators lack are skills to analyze student data and use those data to select and implement interventions effectively with students. Lacking skill, they will experience significant anxiety that may be displayed in the form of resistance (Sailor 2009). Previous research has documented the lack of preservice and in-service preparation of school psychologists in implementing evidence-based practice (White et al. 2007). Moreover, a recent survey conducted of master's level elementary, secondary, and special education teachers found that teachers reported the least amount of preparedness in academic assessment strategies and instructional programs (Begeny and Martens 2006). While special education teachers reported receiving more training in academic assessment than their regular education counterparts, their training was still limited and typically not focused on using data in a problem-solving process. Therefore, professional learning is of paramount concern in ensuring sustained implementation (Sailor 2009).

Previous research has confirmed that teachers who receive training were more likely to (a) implement the program (McCormick et al. 1995), (b) implement programs with integrity (Sanetti and Kratochwill 2009) and (c) report more favorable student outcomes (Simmons 2012). Furthermore, follow up coaching and support is needed for teachers to utilize newly learned skills and maintain those skills. Strong professional learning needs to be the cornerstone of any successful RTI/MTSS implementation.

There is no one model of professional development that has emerged to represent a validated standard of professional learning practice (Kratochwill et al. 2007). However, national standards do exist to guide in designing and delivering professional learning. Learning forward (www.learningforward.org), with the collaboration of 40 professional associations, has developed *Standards for Professional Learning* (2011). The third iteration of these standards outline the characteristics of professional learning that lead to effective teaching practices, supportive leadership, and improved student results. Educators are to take an active role in using data to determine professional learning needs and embed professional learning experiences in the contexts in which the knowledge and skills are to be applied.

Research has identified that job-embedded professional learning structures such as teacher networks and/or study groups were more effective in changing teacher practice than traditional workshop or conference formats and were strongest when groups of teachers were from the same school, department, and/or grade level (Porter et al. 2000). In addition, building opportunities within professional learning experiences for active engagement of adults, content focus, and coherence result in greater changes in practice, and the effect is even stronger when professional learning activities are directly aligned with teacher goals and state standards. Brown-Chidsey and Steege (2005) emphasized three essential elements pertaining to training teachers to implement an RTI/MTSS framework: schedule, teacher learning outcomes, and indicators of mastery of RTI/MTSS methods. They recommended scheduling several training sessions for educators rather than "one shot workshops" or multiday conferences. Educators need ongoing support and training to maintain a high degree of implementation. Learning outcomes should be clearly stated for each training session, and some type of measure of implementation integrity should be used to determine whether teachers are applying newly learned skills in their instruction and decision-making. Finally, just as we differentiate instruction for students, professional development should be differentiated according to teacher needs. A sample training plan is included in Appendix B.

Incentives

Providing incentives to promote implementation of an RTI/MTSS framework is a necessary implementation component. Without incentives, districts and buildings will experience slower change (Hess 2013ss). While many people think incentives need to involve financial resources, many nonfinancial incentives may be provided including giving permission to abandon ineffective existing practices, using data to celebrate incremental success, and providing public praise and recognition to grade level teams who increase student proficiency toward district targets. In addition, some building principals treat membership on a building RTI/MTSS team the same as other instructional duties required of teachers such as lunch room, playground, and/or bus supervision. School leaders can reinforce participation on RTI/MTSS teams by removing another duty from the teacher or staff person's workload.

Of course, there are financial incentives that can be leveraged to speed up the implementation process. Some districts provide stipends to teachers to facilitate grade-level team meetings or professional learning communities. Grade-level team facilitators are reimbursed for the extra time they spend in training, gathering, and generating data reports, and working with their colleagues on improving instructional practices at the grade level. Another financial incentive is to offer "mini-grants" either at the state or district level to work on a specific area of implementation. For example, some larger districts offer mini-grants that teachers, buildings, or grade levels can apply for to help align curriculum and instruction to standards, develop standard treatment protocol interventions, and/or develop common formative assessments. Finally, Federal initiatives such as the *race to the top* (RTTT) grant competition have intensified accountability and prompted new attention to the processes used for the evaluation of teachers and school administrators (Kane et al. 2011). Many states have designed "Pay for Performance" systems where teachers are financially rewarded for growth in student achievement rather than the tradition compensation system where teaches received pay raises for years of teaching and accruing credits for professional development and credits past the bachelor's degree level (Hess 2013).

Measuring Implementation at the State and District Level

The weak link in implementing RTI/MTSS is incorporating routine and reoccurring feedback from implementation measures to those responsible. Implementation integrity has frequently been mentioned as an important component of conducting interventions with students (Gresham 1989; Roach and Elliott 2008; Sanetti and Kratochwill 2009), but measuring implementation the RTI/MTSS system and processes is fundamentally different.

Measuring State-Level Implementation

The role of states in the implementation of RTI/MTSS has been defined as policy development and capacity building (Batsche et al. 2005; Batsche et al. 2007; Callender 2009; Sailor 2009). Only districts through their buildings can actually implement RTI/MTSS. The National Center on Response to Intervention (NCRTI) (http://www.rti4success.org) was funded for 5 years in 2008 to support states in the implementation of RTI/MTSS. NCRTI's website tracks state-level implementation and shares information on state-level policies and implementation activities. Given that no explicit national methodology for RTI/MTSS has been promulgated, state implementation and its measurement has been more informal. No systematic measures of state-level implementation exist. Consequently, Zirkel (2011) notes implementation of RTI/MTSS has been uneven across states, often reflecting the differing needs of states and their political cultures.

Measuring District-Level Implementation

With the inception of RTI/MTSS as an approach in schools, district-level fidelity of implementation measures have been recommended (Mellard and McKnight 2006) and more than 60 examples of implementation measures can be accessed at the TIERS Group website (http://www.hdc.lsuhsc.edu/tiers/Resources.php). Systematically measuring implementation of RTI/MTSS is especially critical to sustainability given the reactions of some educators to implementation who contend that implementation raises too many concerns (Reynolds and Shaywitz 2009; Samuels et al. 2011; Wiener and Soodak 2008). Accurate measurement of implementation ensures that districts have actionable information about adherence to planned change.

District Implementation at SCRED

One example of district-level implementation measurement is the St. Croix River Education District (SCRED), a group of six school districts in East Central Minnesota, who have been implementing the RTI/MTSS Framework formally since 1995. SCRED has a long history of using data-based decision-making through a problem-solving model. Since 1995, the focus has been on working with member districts to implement an RTI/MTSS model that coordinates three critical elements: (a) frequent and continuous measurement using general outcome measures of achievement, (b) research-based instructional practices within a tiered service delivery model, and (c) school-wide organization principles to ensure the most effective instruction for each student. In addition, SCRED districts use data from the RTI/MTSS framework as part of the evaluation process for special education services in the category of specific learning disability.

Measurement Tools for Implementation

To sustain implementation, SCRED has embraced a model where the RTI/MTSS framework is evaluated through its implementation. There is not one global measure that is used to evaluate implementation. Rather multiple sources of data are used, and we look for converging data to identify areas that need more support or refinement. The primary data sources that are used to evaluate implementation efforts are general outcome measures of student achievement (i.e., Aimsweb, Pearson, 2010; Northwest Evaluation Association Measures of Academic Progress) and behavior (i.e., School wide Information System and Aimsweb Behavior), and an RTI/MTSS Implementation Status Checklist developed at SCRED (www.scred.k12.mn.us). All academic and behavior data are stored in a data repository warehouse (Technology and Information Education Services 2013) where data may be aggregated and disaggregated with numerous data report options. Achievement data are used to evaluate the quality of core (tier 1), supplemental (tier 2), and Intensive (tier 3) instruction. In addition, these data are used to track whether the achievement gap is closing for various NCLB subgroups.

Along with achievement and behavioral data, the RTI Implementation Status Checklist (RTI-ISC) is administered on a yearly basis to building teams. The RTI-ISC is divided into six sections: (1) parent involvement, (2) school climate and culture, (3) curriculum and instruction, (4) measurement and assessment, (5) collaborative teams, and (6) problem-solving process. Each section has a series of statements that are evaluated on a continuum of “Not in place”, “limited practice,” “partially implemented,” “mostly implemented”, “well-established”, and “don’t know.” A copy of the RTI-ISC may be found in Appendix C.

In 2010–2011, problem-solving team members and administrators completed the checklist. This group of individuals was most knowledgeable of RTI/MTSS and generally took leading roles in their buildings’ implementation of the framework. However, in 2012, the survey pool was expanded to include other specialists and general education teachers to reflect the multiple tiers of service within an RTI/MTSS framework. One result of this expansion to a wider group was the increased number of “Don’t Know” responses on individual items, resulting in lower average scores. In no way can one conclude that a “Don’t Know” response reflects a deficiency in practice, but simply that the respondent lacks information or knowledge of the practice. As different individuals and teams vary in their responsibilities, individual educators may not be fully aware of the practices of other individuals and teams. For example, a general education teacher may not know that their building has a well-established practice of evaluating assessment and scoring fidelity, and rate the item lower than actual practice. Because these ratings are based on individual perceptions, it is possible that those who critically examine a well-built framework and self-evaluate their processes on an ongoing basis, may actually rate their implementation lower than expected.

Effects of Implementation

Student achievement When SCRED began implementing the RTI/MTSS framework in 1995, the range of student proficiency in reading across grades Kindergarten to 6th grade was 20–38 % on a 1-minute timed reading passage. In 2011–2012, the range of student proficiency was 65–78 % on these same measures. Similar trends have been noted in the area of mathematics. Because SCRED correlates performance on general outcome measures to the Minnesota state accountability test to obtain proficiency cut scores, it is important to note that the proficiency cut points have changed three times since 1995. Each time the state releases a new accountability test, the correlations are recalculated resulting in higher proficiency cut scores.

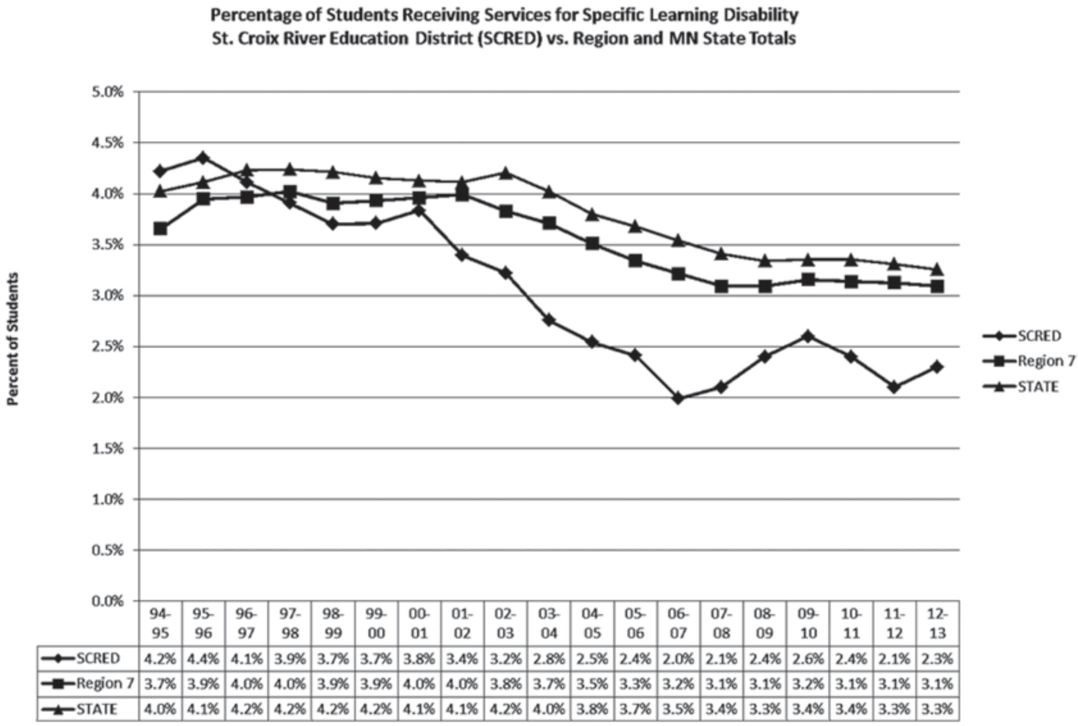
Over the same time period, the district has been tracking the performance of students performing at the 10th percentile. The results indicate that at every grade level, student growth rates in reading have at least doubled and in some instances quadrupled. For example, the median score of second graders at the 10th percentile was 15 words correctly in 1 min in 1996 but had risen to 75 words in 2011. Finally, the percentage of students identified as learning disabled has dropped dramatically over the past decade, by 50 % (See Fig. 3).

It is believed that these data trends provide strong evidence of the preventive nature of the RTI/MTSS framework. Moreover, with the implementation of a multi-tiered service delivery model, teachers realize they are able to get effective interventions in place for students without having to request an evaluation for special education services.

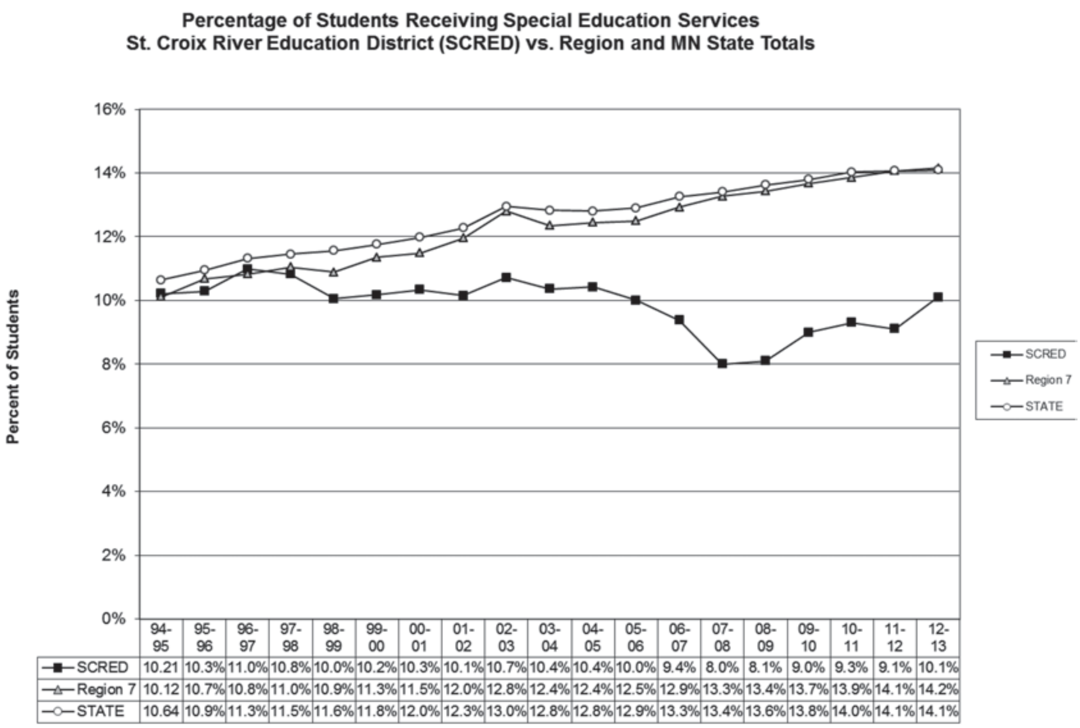
RTI-ISC Average scores across the six different sections of the RTI-ISC for the years 2010–2011 and 2011–2012 are displayed in Table 3. Regarding scores, an average score of "1.0" would indicate that all respondents rated an item as "Don't Know", while a perfect score of "6.0" would indicate that all respondents rated an item as "Well Established." Average SCRED-wide section scores for the 2011–2012 school year fell between 4.06 and 4.8, which could be considered within the "Partially Implemented" to "Mostly Implemented" range. Similarly, average scores across the six sections in 2010–2011 ranged from 4.23 to 4.98. "Don't know" responses on individual items this year ranged from 2.1 to 31.1 %, while the percentage of "Don't Know" responses on individual items in 2010–2011 ranged from 0 % to a high of 22 %. A review of SCRED-wide average scores revealed relative strengths with building/administrative leadership's commitment to the RTI framework, implementing school-wide and individual student assessment plans, informing parents of their child's performance on school-wide assessments, and using results to identify at-risk students and evaluate their progress toward goals. Most buildings obtained their lowest average score on the parent involvement section. Subsequently, parent involvement has been prioritized as a priority area for improvement.

Using Data to Drive Implementation

When the RTI/MTSS framework was first implemented at SCRED in 1995, the student achievement data indicated that there were large numbers of students below proficiency in the area of reading and mathematics. The authors decided to begin focusing on reading, and tackled improving core instruction as a priority. Because there were only between 20 and 38 % of students proficient, there were not enough resources to provide supplemental and intensive instruction to large numbers of students. As students began to make increased progress, developing the tier 2 and tier 3 services began. In 2002, the authors began prioritizing mathematics as a focus area and used the data to guide implementation efforts. Again, core instruction was prioritized over tier 2 and tier 3; however, work was done on expanding the range of interventions available to students. Over the years, new core has been implemented, supplemental, and intensive curricula in reading and mathematics. In addition, the authors have invested heavily in professional development and coaching to assist staff in learning and implementing new skills. Screening and progress monitoring data are used to evaluate the effectiveness of efforts.



	94-95	95-96	96-97	97-98	98-99	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13
SCRED	4.2%	4.4%	4.1%	3.9%	3.7%	3.7%	3.8%	3.4%	3.2%	2.8%	2.5%	2.4%	2.0%	2.1%	2.4%	2.6%	2.4%	2.1%	2.3%
Region 7	3.7%	3.9%	4.0%	4.0%	3.9%	3.9%	4.0%	4.0%	3.8%	3.7%	3.5%	3.3%	3.2%	3.1%	3.1%	3.2%	3.1%	3.1%	3.1%
STATE	4.0%	4.1%	4.2%	4.2%	4.2%	4.2%	4.1%	4.1%	4.2%	4.0%	3.8%	3.7%	3.5%	3.4%	3.3%	3.4%	3.4%	3.3%	3.3%



	94-95	95-96	96-97	97-98	98-99	99-00	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13
SCRED	10.21	10.3%	11.0%	10.8%	10.0%	10.2%	10.3%	10.1%	10.7%	10.4%	10.4%	10.0%	9.4%	8.0%	8.1%	9.0%	9.3%	9.1%	10.1%
Region 7	10.12	10.7%	10.8%	11.0%	10.9%	11.3%	11.5%	12.0%	12.8%	12.4%	12.4%	12.5%	12.9%	13.3%	13.4%	13.7%	13.9%	14.1%	14.2%
STATE	10.64	10.9%	11.3%	11.5%	11.6%	11.8%	12.0%	12.3%	13.0%	12.8%	12.8%	12.9%	13.3%	13.4%	13.6%	13.8%	14.0%	14.1%	14.1%

Fig. 3 Percentage of students receiving services for specific learning disability St. croix river education district (SCRED) versus region and MN state totals

Table 3 Response to intervention implementation status

	Parent Involvement		School climate and culture		Curriculum and instruction		Assessment and measurement		Collaborative teams		Problem-solving teams	
	2011	2012	2011	2012	2011	2012	2011	2012	2011	2012	2011	2012
SCRED	4.23	4.06	4.71	4.7	4.68	4.8	4.98	4.77	4.85	4.64	4.96	4.62
District 1 high school	4.93	4.11	4.93	4.47	5	5.01	5	4.78	4.67	4.75	5.39	4.79
District 1 middle school	4.68	4.49	5.07	5.28	5	5.07	5.37	5.37	5.17	5.31	5.62	5.57
District 1 intermediate	4.12	4.3	4.61	5.07	3.9	4.58	5.1	4.74	4.97	4.38	4.77	4.76
District 1 elementary	4.95	4.87	5	5.18	4.96	5.29	5.38	5.3	5.27	5.25	5.42	5.24
District 1 elementary	4.67	4.5	5.2	5.15	5.12	4.9	5.56	5.46	5.69	5.59	5.79	5.74
District 2 high school	3.34	3.03	3.97	3.38	4.32	3.84	4.22	4.02	4.39	3.31	4.32	3.55
District 2 elementary	3.04	4.34	3.66	4.57	4.09	4.86	4.14	4.83	3.91	4.36	3.69	4.31
District 3 high school	4.28	4.67	4.98	5.14	4.85	5.01	4.91	5.26	5.24	5.38	4.96	5.43
District 3 elementary	4.3	4.5	4.89	5.32	5.4	5.02	5.65	4.93	5.48	5.49	5.47	5.15
District 3 elementary	4.13	4.73	5.69	5.67	5.4	5.62	5.82	5.56	5.62	5.87	5.73	5.73
District 4 high school	3.4	2.28	4.61	3.89	2.33	3.81	3	2.81	2.9	2.29	2.13	1.88
District 4 middle school	2.85	4	3.64	4.58	3.43	4.59	3.58	4.48	3.27	4.34	3.2	3.96
District 4 elementary	3.9	4.46	4.62	4.89	3.9	4.98	4.52	5.11	3.37	4.84	2.9	4.47
District 5 high school	4.44	3.61	4.64	4.2	4.8	4.77	5.37	4.47	5.08	4.53	5.55	4.56
District 5 elementary l	4.6	4.51	4.92	5.12	4.93	5.24	5.05	5.33	4.93	5.19	5.29	5.05

Table 3 (continued)

	Parent Involvement		School climate and culture		Curriculum and instruction		Assessment and measurement		Collaborative teams		Problem-solving teams	
	2011	2012	2011	2012	2011	2012	2011	2012	2011	2012	2011	2012
District 6 high school	4.4	3.38	3.74	3.82	4.49	3.53	4.56	3.8	4.61	4.51	5.47	4.96
District 6 elementary	4.4	3.54	5.19	5.16	5.03	5.14	5.36	4.99	5.15	5.03	5.79	5.29

Implications for Practice

Our own implementation experiences along with a review of common implementation barriers across the country have resulted in many lessons learned and implications for practice. The top ten lessons learned are described in the following sections.

Lesson #1: Understand the Magnitude of System Change

As has been discussed in this chapter, system change is incredibly complex. While it is imperative that districts have a sense of urgency toward improving student outcomes, implementation will likely fail if the process moves too fast without enough support. Districts should expect questions and occasional conflict. Anytime a system is undergoing change, conflict should be expected. It is important for staff to focus on data and student outcomes.

Lesson #2: Begin with a Leadership Facilitated, Clearly Written Implementation Plan That Articulates the District's Vision

In the authors' experiences, many districts attend a workshop or two on RTI and then dive into implementation without a formal implementation plan. When you couple the lack of a plan with high levels of administrative turnover across the country, the end result is many false starts leading to increased resistance among staff.

Lesson #3: A Multi-Tiered Service Delivery Model Is Critical

RTI/MTSS implementation is predicated on the notion that a continuum of service delivery options is available in each building. In buildings where special education services is the only way of providing help to students, large numbers of special education referrals can be expected. Developing a multi-tiered model requires efficient use of resources and a great deal of instructional "teaming." Start with tier 1.

It is recommended that districts examine benchmark data and work on improving core instruction if large numbers of students are not proficient. Teams should then design standard treatment protocol interventions for groups of students as a starting point. Of course, the principal will need to build a master schedule that supports supplemental interventions. If students do not make progress with a standard protocol intervention, then problem-solving teams can customize individual interventions using the problem-solving process. Teams can become easily overwhelmed if they try to individualize interventions for all students rather than start with group level interventions. In fact, if your first step in implementing an RTI/MTSS framework is to start a problem-solving team, then your RTI/MTSS model will likely fail.

Lesson #4: Principals Must Be Instructional Leaders

As instructional leaders in their building, it is imperative to involve principals/school leaders early and often in the RTI/MTSS planning and implementation process. Principals should be active participants on the building problem-solving team to assist in providing support and resources for at-risk students. They should also hold grade-level teams accountable for reviewing grade-level data and evaluating their instructional practices. Finally, they need to build a master schedule to support supplemental and intensive interventions.

Lesson #5: Keep the Focus on Instruction and Provide Ample Support Through Coaching

High-quality professional development should be provided to teachers on improving their instructional practices with research-based strategies. To sustain RTI/MTSS implementation, districts are advised to build in an active "coaching" component to assist staff in skill development. Ongoing training should be provided to grade-level teams and problem-solving teams on data-based decision-making and research-based interventions. Districts are advised to identify an in-district expert "troubleshooter" to assist buildings with RTI/MTSS implementation. Avoid the pitfall of thinking that the first step is to purchase published intervention programs.

Lesson #6: Place As Much Emphasis on Training Teachers How to Use Data As on How to Collect Data

While it is important to train staff on proper administration and scoring procedures for screening, diagnostic, and progress monitoring measures, it is as important to provide training and support on how to use the data to guide instruction. Teachers should be taught to never make an important education decision about a student based on a single data point. Rather, they should be trained to look for converging data that support a hypothesis.

Lesson #7: Ongoing Professional Learning and Support Is Critical

Districts should not confuse awareness training with implementation training. Awareness training is an important component in building consensus; however, staff will need training and follow up support on components of the RTI/MTSS framework. Professional learning should be aligned to needs identified through implementation efforts.

RTI/MTSS implementation is a work in progress. It is important that all key stakeholders recognize that the system needs to be evaluated regularly and modifications and adjustments made accordingly. Decisions cannot be made in a top-down manner. All key stakeholders need to work together to implement the process and address questions.

Lesson #8: Documentation of the Process Is Critical

While RTI/MTSS is primarily about system reform of regular education, there is a special education entitlement component. Accordingly, schools must have a clearly defined problem-solving process with forms and guidelines to guide the process. SCRED found that even with intensive oversight of the process, problems with documentation still existed.

Lesson #9: Develop a System to Measure Intervention Integrity

Within an RTI/MTSS model, it is critical to evaluate whether interventions are implemented as designed. It is important to highlight that this process is meant to be supportive rather than evaluative. Schools are encouraged to discuss in advance a process for determining which team members will conduct integrity observations and what the process will look like.

Lesson #10: Avoid Entitlement Decisions Until Core RTI/MTSS Features Are in Place

Districts will have great difficulty making entitlement decisions if all elements of RTI/MTSS are not in place. For example, if a formative assessment system is not in place, districts will be unable to assess a student's rate of progress. If problem-solving teams do not use a systematic decision-making model, interventions will not be individualized. Moreover, if the team does not use scientific, research-based interventions, student progress is likely to be minimal. Thus, using RTI/MTSS data to make LD identification decisions before the district can document that (a) they are using an well-established RTI/MTSS model, (b) it is being implementing as intended, (c) the interventions are occurring with fidelity, and (d) the problem-solving process is used, will likely result in negative outcomes for students.

Future Directions for Research

In reality, RTI/MTSS can be very complex both in theory and in practice. Although much advancement has been made in intervention, assessments, and decision-making, there is still much work to be done. Future research should focus on several areas. First, fidelity of implementation of the framework at the system level should be researched to determine the most critical features that must be in place and which features are unnecessary. Second, a cost/benefit analysis could inform state and district-level policy-makers. Third, as more research is conducted, meta-analyses to obtain effect sizes for essential components of the RTI/MTSS framework would be beneficial to compare the approach to others that have been attempted. Finally, secondary school settings represent unique challenges exacerbated by the implementation of common core state standards. More descriptive research is needed to explore variations among settings and states.

Summary

This chapter discussed how educators and systems can institutionalize RTI/MTSS in ways that preserve positive changes and increase sustainability over time (until your retirement). School districts around the country have ample and typically uneven experience in implementing new initiatives. However, often times, initiatives are not always interwoven or connected. This chapter was about how to implement and sustain RTI/MTSS at the system level, not with individual students. At this point, the largest challenge is to implement frameworks that result in positive outcomes for students and sustain these frameworks over time. Ultimately, system implementation of an RTI/MTSS framework is very challenging work requiring measurement and feedback to critical decision-makers. Successful RTI/MTSS systems are the result of disciplined leadership coupled with a great deal of knowledge and persistence in organizational routines. Thankfully, there are many talented educational professionals who are up to the challenge, and embrace the opportunity to build models that are truly responsive to the needs of children and the contexts in which they are served.

References

Allison, E., Besser, L., Campsen, L., Cordova, J., Doubek, B., Gregg, L., Kamm, C., Nielsen, K., Peery, A., Pitchford, B., Rose, A., Ventura, S., & White, M. (2010). *Data teams: The big picture—looking at data teams through a collaborative lens*. Englewood: Lead & Learn Press.

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., Reschly, D., Schrag, J., & Tilly, D. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria: National Association of Directors in Special Education.

Batsche, G., Curtis, M., Dorman, C., Castillo, J., & Porter, L. (2007). The Florida problem-solving/response to intervention model: Implementing a statewide initiative. In S. Jimerson, M. Burns, & A. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Begeny, J. C., & Martens, B. K. (2006). Assessing pre-service teachers' training in empirically-validated behavioral instruction practices. *School Psychology Quarterly, 27,* 262–285.

Brown-Chidsey, R., & Steege, M. W. (2005). *Response-to-intervention: Principles and strategies for effective practice*. New York: Guilford.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response–to-intervention. *School Psychology Quarterly, 22,* 297–313.

Burns, M. K., & Gibbons, K. (2012). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices* (2nd ed.). New York: Routledge.

Callender, W. (2009). The Idaho results-based model: Implementing response to intervention statewide. In S. Jimerson, M. Burns, & A. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Deal, T., & Peterson, K. (1998). *Shaping school culture: The heart of leadership*. San Francisco: Jossey-Bass.

DuFour, R. (2004). Schools as learning communities. *Educational Leadership, 61*(8), 6–11.

DuFour, R., DuFour, R., Eaker, R., & Many, T. (2010). *Learning by doing: A handbook for professional learning communities at work*. Bloomington: Solution Tress Press.

Eric, H., Paul, P., & Ludger, W. (2010). *U. S. math performance in global perspective: How well does each state do at producing high-achieving students?* Boston:

Harvard's Program on Education Policy and Governance Report No. 10–19.
Feldman, M., & Pentland, B. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 94–118.
Ferriter, W., Graham, P., & Wight, M. (2013). *Making teamwork meaningful: Leading progress-driven collaboration in a PLC at work*. Bloomington: Solution Tree Press.
Fixsen, D., Naoom, S., Blaise, K., Friedman, R., & Wallace, F. (2007). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida.
Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.
Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York: Teachers College Press.
Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18,* 37–50.
Hanushek, E., Peterson, P., & Woessmann, L. (2010). U.S. math performance in global perspective: How well does each state do at producing high achieving students? Cambridge: Harvard University. http://files.eric.ed.gov/fulltext/ED513539.pdf. Accessed 21 Oct 2014.
Hess, F. M. (2013). *Cage-busting leadership*. Cambridge: Harvard University Press.
Hoerr, T. (2005). *The art of school leadership*. Alexandria: ASCD.
Honore', R. (2012). *Leadership in the new normal: A short course*. Lafayette: Acadian House.
Horner, R. H., Albin, R. W., Todd, A. W., Newton, J. S., & Sprague, J. R. (2010). Designing and implementing individualized positive behavior support. In M. E. Snell & F. Brown (Eds.), *Instruction of students with severe disabilities* (7th ed.). Upper Saddle River: Pearson Education, Inc.
Jackson, Y., & McDermott, V. (2012). *Aim high, achieve more: How to transform urban schools through fearless leadership*. Alexandria: ASCD.
Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Next, 11,* 1–11.
Knoster, T., & George, H. P. (2002). Realizing durable and systemic behavior change in schools: Guiding questions. *NASP Communique, 30*(6), 34–37.
Kratochwill, T. R., Volpiansky, P., Clements, M., & Ball, C. (2007). Professional development in implementing and sustaining multitier prevention models: Implications for response to intervention. *School Psychology Review, 36,* 618–631.
Levin, H. M., Catlin, D., & Elson, A. (2005). *Costs of implementing adolescent literacy programs*. New York: Carnegie Corporation of New York.
Marzano, R., Waters, T., & McNulty, B. (2005) *School leadership that works: From research to results*. Alexandria: ASCD.
McCormick, L. K., Steckler, A. B., & McLeroy, K. R. (1995). Diffusion of innovations in schools—A study of adoption and implementation of school-based tobacco prevention curricula. *American Journal of Health Promotion, 9,* 210–219.
McNulty, B., & Besser, L. (2011) *Leaders make it happen: An administrator's guide to data teams*. Englewood: Lead & Learn Press.
Mellard, D. F., & McKnight, M. A. (2006). *RTI implementation tool for reading: Best practices* [Brochure]. Lawrence: National Research Center on Learning Disabilities.
Murphy, C. (2006). *The whole-faculty study groups fieldbook: Lessons learned and best practices from classrooms, districts, and schools*. New York: Corwin.
National Institutes of Health. (2013). *Frequently asked questions about implementation science*. http://www.fic.nih.gov/News/Events/implementation-science/Pages/faqs.aspx. Accessed 19 March 2013.
Payne, C. (2008). *So much reform so little change: The persistence of failure in urban schools*. Cambridge: Harvard Education Press.
Porter, A. C., Garet, M. S., Desimone, L., Yoon, K. S., & Bierman, B. F. (2000). *Does professional development change teaching practice? Results from a three-year study*. Washington, DC: American Institutes for Research in Behavioral Sciences.
Project Forum. (2007). *Response to intervention as it relates to early intervening services: Recommendations*. Alexandria: National Association of State Directors of Special Education.
Reeves, D. (2005). *Accountability in action: A blueprint for learning organizations*. Englewood: Lead & Learn Press.
Reynolds, C., & Shaywitz, S. (2009). Response to intervention: Ready or not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly, 24*(2) 130–145.
Roach, A., & Elliott, S. (2008). Best practices in facilitating and evaluating implementation integrity. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (pp. 195–208). Bethesda: National Association of School Psychologists.
Sailor, W. (2009). *Making RtI work: How smart schools are reforming education through schoolwide response-to-intervention*. San Francisco: Jossey Bass.
Samuels, C. A., Sawchuk, S., Zehr, M. A., Kelleher, M., & Sparks, S. D. (2011, March 2). Special report: Response to intervention's promise and pitfalls. *Education Week*, pp. S1–S16.
Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445–459.
Sarason, S. (1996). *Revisiting the culture of the school and the problem of change*. New York: Teachers College Press.
Seashore, L. K., Wahlstrom, K., & Anderson, S. E. (2010). *Learning from leadership: Investigating the links to improved student learning*. Final report of Research to

The Wallace Foundation. University of Minnesota and University of Toronto.

Senge, P. M. (1990). *The fifth discipline: The arts and practice of the learning organization*. New York: Currency/Doubleday.

Shelton, S. (2011). *Strong leaders strong schools: 2010 school leadership laws*. New York: Wallace Foundation.

Sherer, J., & Spillane, J. (2011). Constancy and change in work practice in schools: The role of organizational routines. *Teachers College Review, 113*(3), 611–657.

Simmons, W. (2012). Data as a lever for improving instruction and student achievement. *Teachers College Record, 114,* 1–9.

Spectrum K12 School Solutions. (2010). Response to Intervention Adoption Study. http://rti.pearsoned.com/docs/RTIsite/2010RTIAdoptionSurveyReport.pdf.

Technology and Information Education Services. (2013). TIES Data System [Computer software]. St. Paul, MN: TIES. http://www.ties.k12.mn.us.

The Wallace Foundation. (2013). The school principal as leader: Guiding schools to better teaching and learning.

Ward, R. (2004). Defining the school's culture. In R. Ward & M. Burke (Eds.), *Improving achievement in low performing schools: Key results for school leaders.* Thousand Oaks: Sage.

Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21,* 1–9.

White, J., Kumpke, P., & Kratochwill, T. R. (2007). *School psychologists' use and awareness of treatment guidelines: Present and future*. Manuscript submitted for publication.

Wiener, R. M., & Soodak, L. C. (2008). Special education administrators' perspectives on response to intervention. *Journal of Special Education Leadership, 21,* 39–45.

Zamuda, A., Kuklis, R., & Kline, E., (2004). *Transforming schools: Creating a culture of continuous improvement*. Alexandria: ASCD.

Zirkel, P. (2011). Professional practice: State laws and guidelines for RtI: Additional implementation features. *NASP Communique, 39*(7), 21–22.

Evaluating the Impact of Response to Intervention in Reading at the Elementary Level Across the State of Pennsylvania

Edward S. Shapiro

The underlying key element in evaluating the implementation of a response-to-intervention (RTI) process in schools is whether overall student achievement has improved. The simple question of whether RTI is working cannot be answered with an equally simple "yes" or "no" response, but requires examination of multiple questions that collectively allow the implementer to know whether or not the model is accomplishing its desired goals.

Although past efforts to evaluate RTI implementation have reported outcomes on specific metrics such as the number of students referred for special education evaluation and the number of special education placements (e.g., Callendar 2007; Hartman and Fay 1996; Marston et al. 2003; VanDerHeyden et al. 2007), the number of students that responded to the intervention or did not respond and required additional or more intensive intervention (e.g., Ardoin et al. 2005; McMaster et al. 2005; Vaughn et al. 2003; Vaughn et al. 2007), or change in student achievement levels or growth in academic skills following the implementation of an RTI model (e.g. Fuchs et al. 2004; Gettinger and Stoiber 2008; Marston et al. 2003; McMaster et al. 2005), any one of these metrics alone would be insufficient to know whether the implementation was meeting its objectives. Indeed, Peterson et al. (2007) in their description of an evaluation of an RTI service delivery model being implemented in Illinois considered multiple questions related to student outcomes, resource allocation, referral rates to special education, and parental satisfaction. Likewise, Bollman et al. (2007) in their evaluation of a model of RTI identified key indicators over a 10-year period such as increasing percentage of students at benchmark, increasing percentages of students meeting criteria on statewide assessments, and reductions in students identified as having specific learning disabilities (SLDs) across time. Clearly, the evaluation of an RTI model requires a conceptual framework that includes multiple perspectives and measures to determine the overall effect.

Shapiro and Clemens (2009) identified a broad conceptual framework for examining the outcomes of RTI that focused on student achievement at the individual and system level. In particular, when the data in response to five questions posed by Shapiro and Clemens (2009) are aggregated across schools, one can effectively examine the impact of larger questions of evaluation at district and state levels. Specifically, the five evaluation questions asked in the Shapiro and Clemens (2009) framework examined: (1) changes in percentage of student risk level across benchmarks assessments, (2) changes in rate of improvement (ROI) across benchmark assessment, (3) movement between tiers between benchmark periods, (4) movement within tiers between benchmark periods, and (5) percentage

E. S. Shapiro (✉)
Center for Promoting Research to Practice, Lehigh University, Bethlehem, USA
e-mail: ed.shapiro@lehigh.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_37

of students referred for special education found to be eligible. When these questions are addressed simultaneously, one obtains a holistic perspective on the impact of an RTI model.

The purpose of this chapter is to describe the evaluation of a statewide model of RTI using a modification of the Shapiro and Clemens (2009) framework. Specifically, the chapter first presents the Pennsylvania model of RTI, a model that was developed and implemented beginning in 2005 and is recognized as the accepted framework for schools desiring to implement RTI within Pennsylvania. Next, the description of an evaluation of a subset of self-selected schools across a year of implementing the RTI model using the modified Shapiro and Clemens (2009) framework is presented. Finally, implications for practice and research regarding the evaluation process are described.

Pennsylvania's Response to Instruction and Intervention Framework

Pennsylvania developed a statewide model of RTI in 2007 that was a standards-aligned, comprehensive school improvement framework used to enhance outcomes for *all* students within the continuous process of systems-level improvement. Although the model was initially labeled as RTI, the Pennsylvania Department of Education changed the name of the model to response to instruction and intervention (RTII). The change to RTII communicated the message and understanding that core instruction was foundational and that core and supplemental instruction work in tandem within a comprehensive and effective multi-tiered system of support.

RTII has always carried dual meaning in Pennsylvania. Not only does the model represent a viable approach to systems-level improvement but it is also considered an alternate method for the identification of SLDs with a formal process outlined by the Bureau of Special Education within the PA Department of Education. Schools use the RTII/SLD application tool as a reliable implementation checklist that promotes one common approach to implementation, as well as part of a formal approval process to use RTII methodology to identify SLDs. These tools are available on the web site of the Pennsylvania Training and Technical Assistance Network (PATTAN) (http://www.pattan.net).

Figure 1 illustrates the model of RTII in Pennsylvania. Similar to most other RTI models, the

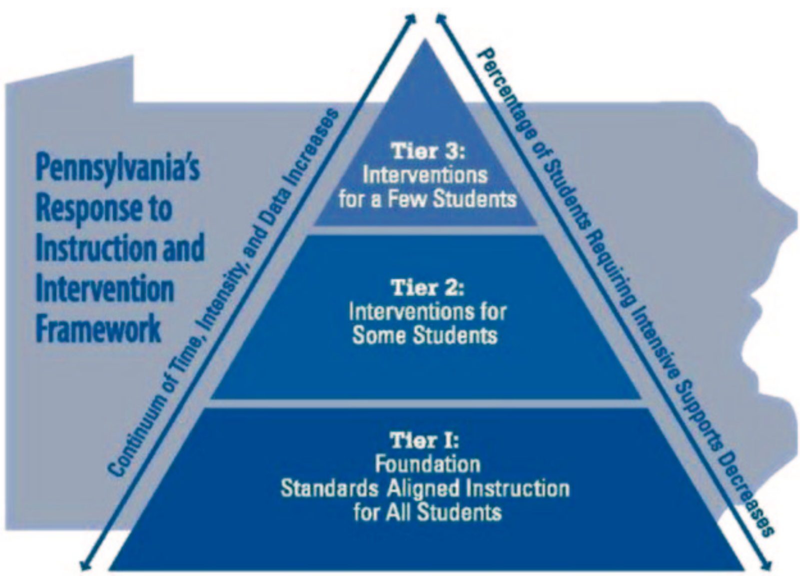


Fig. 1 Pennsylvania response to intervention and instruction model

model in Pennsylvania includes the following core components: (1) standards-aligned core instruction, (2) universal screening, (3) shared ownership of student achievement, (4) data-based decision-making, (5) tiered intervention and service delivery system, and (6) parental engagement. Also similar to models across the country, the heart of RTII in PA is provision of high-quality standards—aligned core instruction to all students. Response to core instruction (tier 1), as measured by the percentage of students who are proficient at both the grade and school level, is used to evaluate the overall health of the general education system. Ongoing evaluation, as measured by reliable and valid systems-level measures, is used to monitor student growth, implement curricular and instructional changes, and identify those students who would benefit from additional support. Characteristics of high-quality core instruction include the use of active learning and engagement strategies, clear and high expectations, differentiated instruction, a continuum of formal and informal assessment, and the use of effective teaching strategies.

Despite exposure to high-quality core instruction, some students require supplemental academic and/or behavioral support. Supplemental instruction at tier 2 is generally provided in a smaller group format several times per week and often targets common deficiencies that exist among the general student population (e.g., phonemic awareness, phonics, vocabulary, mathematics fact fluency, extended core instruction, etc.). Supplemental instruction at this level is provided in 15–30-min increments and student response is monitored at least bi-monthly.

For students who are not responsive to tier-2-level instruction, supplemental instruction that occurs at tier 3 is considered to be the most intensive support that can be offered within the general education system. Support at this level is reserved for students who present with the most significant academic and/or behavioral difficulties. Instructional intensity at this level is often characterized by the smallest teacher to student ratio possible (e.g., 1:3), a minimum of 45–60 min of daily supplemental instruction, frequent progress monitoring, and use of measures that are sensitive to incremental growth. Intensive instruction at this level is highly customized and explicit.

Beginning in 2005, Pennsylvania provided training in developing an RTII model. A total of seven elementary schools were selected to participate in an implementation project focused on reading. Each site established a multi-tiered system of support, goals for improving core instruction, and a continuum of research-based supplemental instruction/intervention methodologies. All sites formed data-analysis teams and developed a schedule for teams to meet regularly and intervene with at-risk students accordingly. All sites administered curriculum-based measures as universal screenings three times per year. Schools also used the 4Sight Benchmark Assessment (http://www.successforall.org) as another universal screening for students in grades 3 and up. The 4Sight Benchmark Assessment is aligned to the statewide annual assessment (Pennsylvania System of School Assessment, PSSA), and has equivalent alternate forms allowing the measure to be given up to four times per year. All sites were provided with ongoing professional development and on-site technical assistance.

Between 2005 and 2009, many other elementary schools chose to implement an RTII model, primarily in reading. All of these schools were provided with technical assistance from the layers of available supports within the state of Pennsylvania. By 2008–2009, RTII was being implemented to some degree within Pennsylvania by a large number of schools.

Evaluation of RTII

Participating Schools

During the 2008–2009 school year, the Pennsylvania Training and Technical Assistance Network (PATTAN) embarked on an evaluation of the outcomes of the implementation of RTII implementation across multiple districts. The evaluation process requested districts across the state to volunteer their data sets for aggregation and analysis. The objective was to examine the student outcome data across schools and districts, to better determine the impact of the Pennsylvania RTII model as it was being implemented across these sites. Because the objective was to obtain data from those sites implementing the model with a reasonable degree of integrity, PATTAN consultants from the eastern, central, and western regions recommended districts/schools that they had been consulting with over the past 2 years toward effective RTII implementation.

PATTAN surveyed the 29 intermediate units (IUs) in the state at the beginning of the 2008–2009 school year and asked them to rate the districts that were implementing RTII as beginning, in process, or experienced implementers of RTII. A total of 11 of the IUs returned this survey. PATTAN staff then contacted all those rated as experienced districts and invited them to be a part of the data pool. A total of 15 districts were sent the initial invitation and data pool survey and nine districts participated. A number of these districts included multiple schools. In addition to the nine districts, those schools in districts where the state RTII pilots had experienced implementation with good integrity of the model over several years were also included, resulting in a total of 15 districts and 29 schools included in the analysis.

Across the 15 districts included in the evaluation, a total of 29 schools were used in the analysis. Schools varied in number of grades included, so the number of schools available for each part of the analysis varied by grade. As indicated in Table 1, the sample for analysis included a wide range of demographic sites from rural and small urban environments, with an equally wide range of minority and socioeconomic representation. All demographics were those reported through the 2006–2007 year on the National Center for Education Statistics (NCES) website (the latest data available at the time the report was completed). Locale designations are those used by the NCES. Also included in Table 1 are the outcomes of the 2008 and 2009 PSSA AYP evaluations for each building.

Framework for the Data Analysis

Consistent with a slight modification of Shapiro and Clemens (2009), the analysis of the RTII implementation was conducted by examining four specific evaluation questions. Responses to each of the following questions served as the framework for the evaluation:

1. What were the identified benchmark risk levels in reading across schools and grades?
   - The analysis was conducted by using each school as the level of analysis, aggregating across grades and across schools.
   - The analysis also examined the *annual yearly progress* outcomes for schools on the PSSA, for those schools with grades 3 through 5.
2. What is the percentage of students that moved across tiers in reading based on identified benchmark scores?
   - The analysis was conducted by using each student as the level of analysis, aggregated across all schools and districts, and reported by grades.
   - Tier changes were determined by outcomes as established on benchmark levels, as determined by curriculum-based measurement, examined fall to winter and then winter to spring.
3. What movement within tiers was evident as reflected in progress monitoring data among students receiving tiered instruction?
   - The analysis was conducted by calculating the yearly attained ROI as well as the targeted ROI (using end-of-year grade-level benchmark scores as the target) for each student for whom progress monitoring was conducted within each grade aggregating across all schools within school districts. Comparisons were also made between the attained ROI of students and the benchmark ROI expected of typical performing grade-level peers who begin and end the year at benchmark, as determined by the dynamic indicators of basic early literacy, 6th edition (DIBELS, Good and Kaminski 2002) data base.

Table 1 Demographics of schools represented in the analysis

District ID	School location	Locale	Total students	% minority	% F/R lunch	Schoolwide title?	Grades	2008 AYP	2009 AYP	2009 PSSA % reading proficient[a]
10001	East	Suburb: large	387	50.1	27.9	Yes	K-6	Yes	Yes	77.1
10002	Central	City: small	289	6.2	39.8	No	K-6	Yes	Yes	77.3
10004	Central	City: small	317	6.3	39.4	Yes	K-6	Yes	Yes	79.1
10005	Central	City: small	478	5.6	58.6	Yes	K-6	SI-1	Progress	74.6
10006	Central	City: small	560	8.8	45.5	No	K-6	Yes	Warning	75.3
10007	Central	City: small	518	3.3	53.9	Yes	K-6	Yes	Warning	66.2
10003	Central	City: small	428	3.3	50.7	Yes	K-6	Yes	Yes	75.4
10008	Central	City: small	419	14.1	84.5	Yes	K-6	Warning	SI-1	51.6
10009	Central	City: small	512	2.9	43.8	No	K-6	Yes	Yes	80.1
10010	Central[b]	City: small	387	10.6	82.2	Yes	K-6	Warning	SI-1	54.2
10011	Central	City: small	340	17.1	88.8	Yes	K-6	Warning	SI-1	52.9
10012	West	Rural: fringe	604	6.1	33.6	No	K-6	Yes	Yes	72.6
10013	East	Suburb: large	516	26.6	53.3	No	K-6	SI-1	Progress	60.3
10014	Central	Suburb: large	382	46.1	33.5	No	K-5	Yes	Yes	65.6
10015	Central	Suburb: large	298	51.3	32.2	No	K-5	Yes	Yes	63.6
10016	Central	Suburb: large	348	54.9	38.2	No	K-5	Yes	Yes	66.1
10017	Central	Suburb:small	452	17.9	18.6	No	K-5	Yes	Yes	72.7
10018	Central	Suburb: midsize	906	10.7	18.3	No	4-6	Yes	Yes	79.8
10019	Central	City: small	473	80.1	94.5	Yes	K-5	Warning	Yes	44.9
10020	Central	Rural: fringe	281	1.8	22.8	No	K-4	Yes	Yes	81.5
10021	West	Suburb: large	425	19.1	38.8	No	K-6 (K-2)	Yes	Yes	80.9
10022	West	Rural:remote	239	<1.0	43.5	Yes	K-5	Yes	Yes	71.2
10023	Central	Rural: distant	541	7	25.7	No	K-5	Yes	Yes	75.2
10024	East	Suburb: large	690	69.3	57.1	Yes	K-5	Progress	Yes	63.3
10025	Central	Suburb: large	608	11.7	13.7	No	K-6	Yes	Yes	75.0
10026	Central	Rural: fringe	580	2.9	9.7	No	K-6	Yes	Yes	82.6
10027	Central	Suburb: large	606	8.1	14.7	No	K-6	Yes	Yes	76.2
10028	Central	Suburb: large	618	7.3	18.6	No	K-6	Warning	Yes	75.9
10029	West	Suburb: large	767	35.7	N/A	Yes	K-5 (K-2)	Cor Ac-2	Cor Ac-2	59.0

AYP adequate yearly progress, *PSSA* Pennsylvania system of school assessment, *F/R* free or reduced price lunch, *SI* school improvement, *Cor Ac-2* corrective action

[a] Proficient = scored in the proficient or advanced range

[b] RTI implemented only in K-2 in this building

4. What were the number of students evaluated for special education SLD eligibility and the number of students found eligible for SLD across districts?
   - The analysis was conducted by examining data reported by school and aggregated by district, if multiple schools are included for the district in the analysis.
   - A limited number of schools provided these data.

Results of the Evaluation

Risk Levels for Reading Performance: Outcomes of Universal Screening

Data were provided for early literacy measures and oral reading fluency (ORF) for 24–29 schools at grades K, 1, and 2, and for 12–15 schools at grades 3–5 (some schools did not collect all early literacy measures for each grade). For grades 3, 4, and 5, 4Sight Benchmark Assessment data were provided from 10 or 11 schools (one school did not collect all 4Sight Benchmark Assessment measures). Outcomes for risk levels were averaged across schools.

Outcomes for the spring for grades kindergarten through grade 5 are shown in Fig. 2. End-of-year performance for students reached its highest level at the end of kindergarten where 89 % of students across 26 schools were found to be at *low risk* on phoneme segmentation fluency (PSF), the measure considered the index measure for end-of-year kindergarten performance. Likewise, an average of 75 % of students reached *low-risk* performance on the secondary kindergarten measure of nonsense word fluency (NWF). Another strong indicator of the very successful outcomes at kindergarten was that the average level of *at risk* students at spring on the PSF measure was 2 % across the 25 schools in the analysis. Across schools at kindergarten, spring PSF low-risk level ranged from 54 to 100 %, with 16 of 26 schools showing low risk of 90 % or better and 23 of 26 schools showing low-risk levels of 70 % or better (see Fig. 3).

At grade 1, spring ORF and NWF measures are considered the index skills reflective of reading performance, with ORF being considered the most important. As shown in Fig. 2, across 28 schools included in the analysis, an average of 73 % of students in spring was found to be at *low*

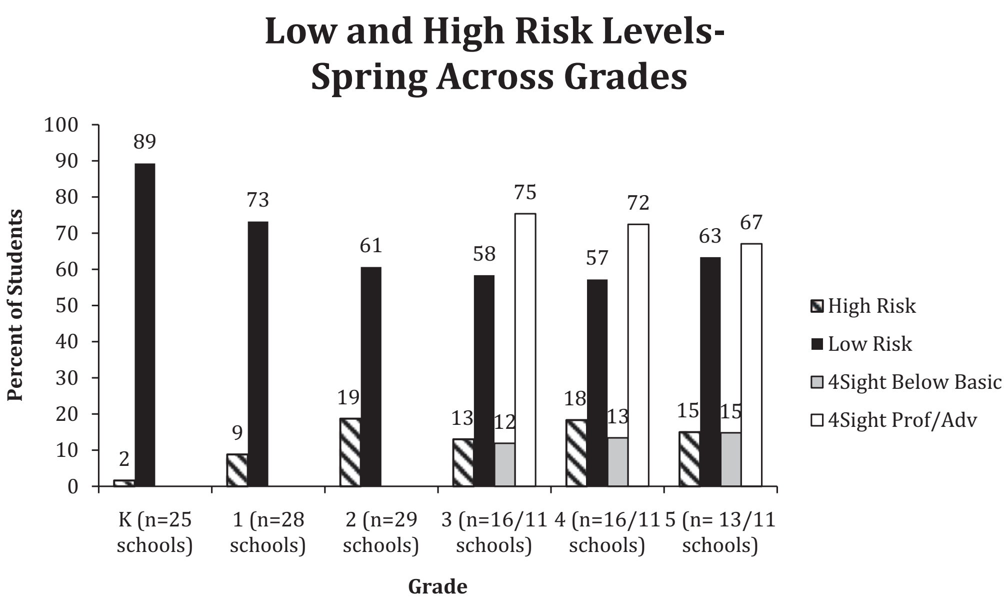


Fig. 2 Risk levels across grades according to Dynamic Indicators of Basic Early Literacy Skills and 4Sight Benchmark Assessments

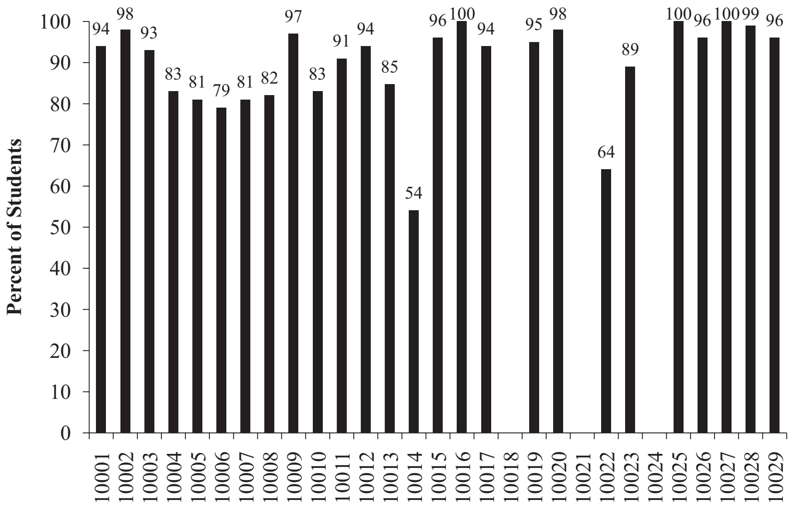


Fig. 3 Percentage of kindergarten students at each school who scored in the low-risk range on the phoneme segmentation fluency benchmark assessment

risk on ORF, and an average of 9 % was found to be at *high risk***.** Across schools at grade 1, spring ORF low-risk level ranged from 43 to 90 %, with 12 of 28 schools showing low risk of 80 % or better and 21 of 28 showing low risk of 70 % or better.

At grade 2, ORF is considered the index skill reflective of reading performance. As shown in Fig. 2, across 29 schools included in the analysis, an average of 61 % of students in spring was found to be at low risk. An average of 19 % was found to be at high risk in spring of grade 2. Across schools at grade 2, spring ORF low-risk level ranged from 28 to 82 %, with 15 of 29 schools showing low risk of 60 % or better and 9 showing low risk of 70 % of or better. Similarly, across schools at grade 2, spring ORF high-risk level ranged from 6 to 56 %, with 15 of 29 schools showing high risk less than 20 % and 27 of 29 showing high-risk levels under 30 % of students.

Considering the outcomes for grade K-2, the level of risk was lowest at the end of kindergarten and increased as one reached grade 2. Given that RTII was designed to build strong foundations in basic educational skills upon which subsequent learning is built across grades, it was not surprising that the strongest effects were seen at the youngest grades. These outcomes are consistent with other evaluations of RTI (Clemens et al. 2011). Clearly, there remains a very wide range of risk levels across schools within each grade. The range appears to become greater as one moves from kindergarten through grade 2.

An examination of grades 3, 4, 5 as shown in Fig. 2 showed that the level of low risk reached around 60 % of students based on the ORF measure with high risk averaging between 13 and 17 % of students across 11 to 16 schools. Looking across schools at each grade showed that at each grade the range of students at low risk was between 40 and 80 %, with most schools reporting low risk levels in the 50–60 % range for grades 3 and 4, and between 60 and 70 % for grade 5. High risk levels across schools remained at or below 20 % in most cases.

Because ORF begins to lose its sensitivity to reflect overall reading performance as students move upward through elementary grades (Shapiro et al. 2008), it was important to also examine risk outcomes as reflected on the 4Sight Benchmark Assessment measures, a measure specifically designed to assess overall reading comprehension skills. As evident in Fig. 2 for grades 3, 4, and 5, an average between 63 and 75 % of students reached proficiency on the spring 4Sight measure across the 11 schools reporting 4Sight data. Comparing the outcomes of ORF and 4Sight showed that the ORF measure in grades 3 and 4 underestimated the percentage of students who were proficient or at low risk, while the results at grade 5 were consistent between ORF and 4Sight.

Overall, the results of the risk-level analysis for reading based on outcomes of ORF and 4Sight Benchmark Assessment showed that the strongest outcomes occurred in kindergarten and grade 1, with less strong outcomes in the middle elementary grades (grades 2, 3, and 4). Together these data reflected that the ORF measure was indeed a very good indicator of those likely to not be very successful in meeting expected levels of reading performance. However, the data also showed that ORF was underestimating the levels of proficient reading achievement for grades 3 and 4 in particular, and somewhat for grade 5. It was important that no single measure, such as ORF or 4Sight be used in making predictions of reading performance for students.

An examination of PSSA outcomes across the 29 schools included in the analyses showed that a total of 21 (72.4 %) made AYP in 2009. Across these schools, the average percentage of students at advanced or proficient was 70.0 %, with a range of 44.9–82.6 %. Between 2008 and 2009, 23 of 29 schools (79.3 %) showed maintenance of AYP or improvement toward AYP while 5 schools (17.9 %) declined in their performance and one school (3.4 %) remained in corrective action 2 for a second year.

Movement Across Tiers

Another important way that the outcomes of the RTII model are examined was the amount of movement across tiers that occurred across the year. Students whose reading performance improved to levels that were greater than the required benchmarks were reflected in this evaluation metric. Tier movement was examined between the beginning- (BOY) and middle-of-year (MOY) assessments and then between the middle- and end-of-year (EOY) assessments. Because the majority of tier movement occurred between fall and winter assessments, only those data are reported in this analysis. Specifically, the percentage of students who begin the year at each risk level (intensive (black), strategic (gray), benchmark (white)) are examined to determine whether they moved up (toward benchmark) or down (toward intensive) at each subsequent assessment period based on their attained benchmark scores. Data for this analysis examined the tier movement for individual students, aggregated and averaged across all schools and displayed by grades.

Figure 4 shows those students who were found at the beginning of the year to be at the intensive (at-risk) level. The largest percentage of movement between beginning and middle of the year occurred at kindergarten with an average of 53 % of these students scoring at benchmark and 19 % remaining at the intensive level. Likewise, at grade 1 a total of 58 % of students moved to either strategic (tier 2) or benchmark (tier 1) at the middle of the year. Grades 2, 3, and 5 showed the least amount of movement to the middle of the year with over 70 % of students remaining in the intensive level. Somewhat more movement was evident at Grade 4 where 40 % of the students moved from the intensive level by the middle of the year.

As seen in Fig. 5, among those whose beginning of the year scores were found to be at the strategic (some risk) level, 67 % of those in kindergarten moved to benchmark by the middle of the year while 13 % dropped to the intensive level (at-risk). From first grade upward, movement toward benchmark occurred between 40 and 50 % of students while between 5 and 16 % moved downward to the intensive level.

Figure 6 shows the beginning to middle of the year movement of students beginning the year at benchmark. Most students in grades 2, 4, and 5 maintained their benchmark status, however, between 12 and 23 % of students in grades K, 1, and 3 showed movement from benchmark to either strategic or intensive levels, with the largest changes occurring in grades K and 1.

In examining movement from middle to end of year (figures not shown), the only grade with substantial changes evident was kindergarten where 75 % of students at the intensive level moved to either the strategic or benchmark levels. Between

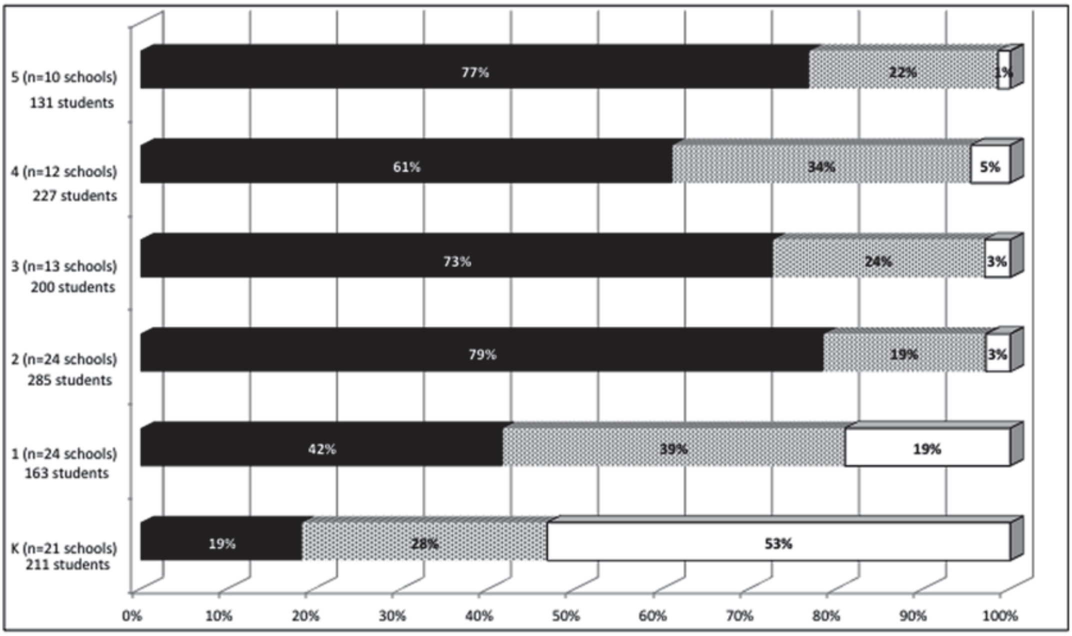


Fig. 4 Tier movement from beginning to middle of the year of students across grades who scored intensive (at-risk, *black*) at the beginning of the year according to Dynamic Indicators of Basic Early Literacy Skills standards

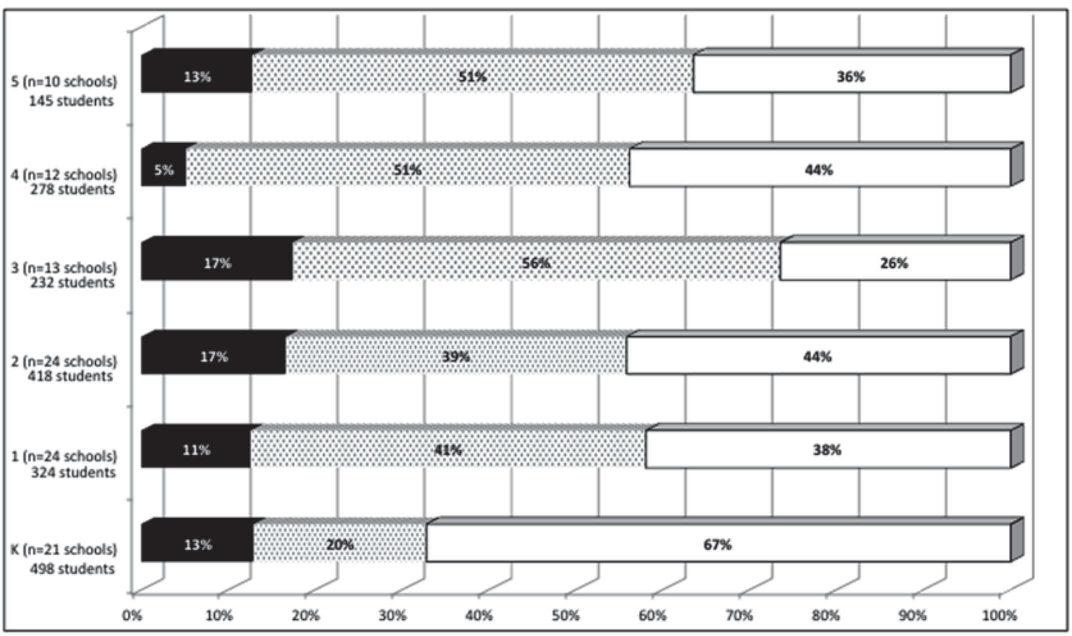


Fig. 5 Tier movement from beginning to middle of the year of students across grades who scored strategic (some risk, *gray*) at the beginning of the year on Dynamic Indicators of Basic Early Literacy Skills

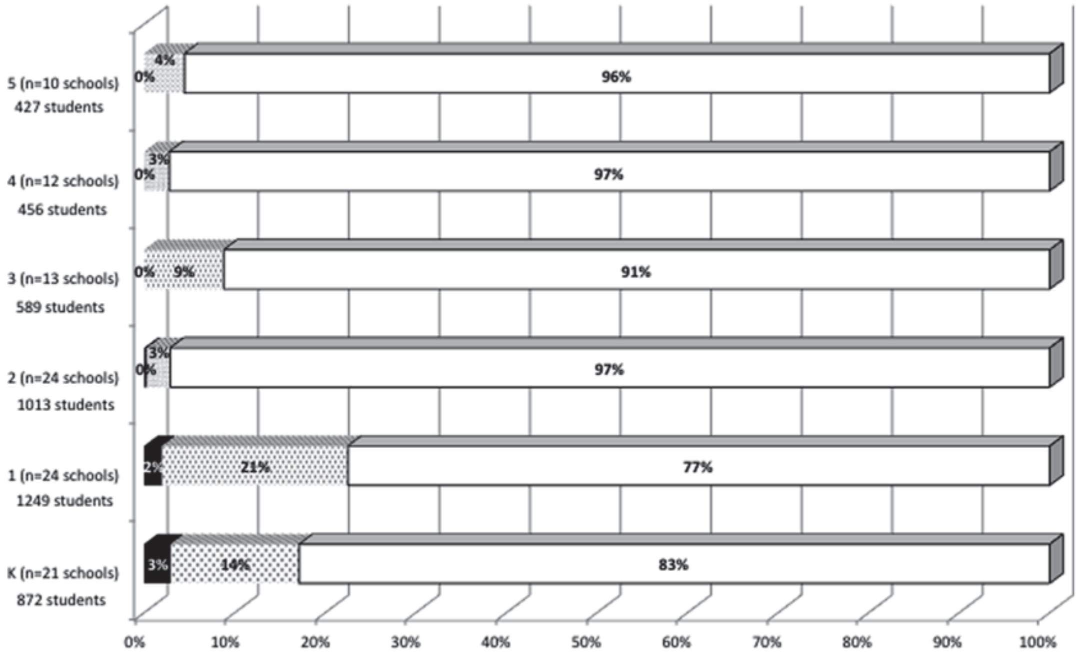


Fig. 6 Tier movement from beginning to middle of the year of students across grades who scored benchmark (low risk, *white*) at the beginning of the year on Dynamic Indicators of Basic Early Literacy Skills

grades 1 and 5, 55–86% of students remained in the intensive level at the end-of-year assessment. A similar pattern was present among those in kindergarten scoring strategic at the middle of the year with 63% of these students moving to benchmark at the end of the year. Movement was also noted among those students scoring strategic in grade 1 at the middle of the year where 42% of students moved upward from strategic to benchmark, and 16% moved downward to intensive. Grade 2 also showed 31% of students moving from strategic to intensive. From grades 3 through 5, between 60 and 74% remained at the strategic level at the end-of-year assessment, with between 12 and 17% moving to intensive and 12 and 28% moving to benchmark. The largest movement toward benchmark occurred at grade 5 where between 85 and 94% of students scoring at benchmark at the middle of the year remained at benchmark at the EOY assessment.

Overall, the movement across tier data indicated that the largest amount of movement occurred in the earliest grades and from the beginning to middle of the year. The data reinforced the importance of early intervention and the need to attend to improving literacy skills at the youngest grades. Although change was still evident at higher grades as well as from the middle to the end of the year, the stabilization of student performance over grades and time once students reached the middle of the school year was the predominant finding from these data.

Movement Within Tiers

The amount of movement occurring within tiers as reflected in student progress monitoring is another indicator of outcomes of the impact of the RTII model. Although students may not make sufficient progress to be moved from one tier to another, those receiving supplemental instruction should be improving in a way that closes the gap between themselves and their peers. One way to capture this movement was to examine and aggregate progress monitoring data across students in ways that reflect student gains against typical performing peers.

Because some of the districts reporting processes did not lend itself to aggregating progress monitoring data, the number of districts and schools for which these data were reported were

Table 2 Total number of students included in progress monitoring data analysis across school districts (schools)

District/school ID	Schools	K	1	2	3	4	5
10001	1	< 10	7	21	21	34	18
10002–10011	10	115	180	264	34	28	N/A
10012	1	40	28	80	57	N/A	N/A
10014–10016	3	77	81	71	64	76	42
10017	1	26	13	23	< 10	3	17
10018	1	N/A	N/A	N/A	N/A	65	10
10020	1	< 10	11	19	30	13	N/A
10021	1	22	17	11	11	< 10	N/A
10024	1	N/A	50	45	67	60	68
10025–10028	3	14	49	72	93	85	68
10029	1	10	32	36	N/A	N/A	N/A

fewer than those for the previous two indicators of risk and movement across tiers. Also, for those schools reporting progress monitoring for less than ten students in a grade, data for that grade from that school was not included in the analysis. Data were averaged within grades across all schools for a single district. Table 2 displays the districts, number of schools, and total number of students per grade for whom progress monitoring data were provided.

The ROI was calculated for each student using the *ordinary least squares* regression (slope) for the progress monitoring data that was implemented during the year. Any students with less than four data points for progress monitoring were excluded. Also, students who were monitored in below grade-level materials were not included in the analysis. Goals for all students were set to the end-of-year grade-level benchmark. Finally, for purposes of the analysis, students who were assigned to strategic or intensive (tiers 2 or 3) interventions were combined. As a result of these assumptions, the analysis represented a conservative estimate of outcomes of movement within tiers as reflected through progress monitoring data.

Figure 7 shows a comparison between the averaged attained ROI and targeted ROI across students by grade. At kindergarten, students averaged an attained ROI that was substantially larger than their targeted ROI. This was also evident at grade 5. At kindergarten, the gap between what students should have attained and what they actually attained was substantially reduced. Reductions of the gap between target and attained outcomes were also slightly evident for grade 5. At grades 1, 2, 3, and 4, the average attained ROI was found to be less than the average targeted ROI.

When compared to the typical ROI expected of similar grade-level students, on average those in grades 1, 2, and 3 were not making as much progress as same grade-level students who began and ended the year at benchmark. At fourth grade, average attained ROI exceeded that expected of typical fourth graders, but did not reach levels of targeted ROI. Finally, at grade 5, student progress monitoring performance showed attained ROI slightly greater than targeted ROI as well as an attained ROI almost twice as large as typical fifth graders who begin and end the year at benchmark.

Given the wide range of starting points for progress monitoring across students, it was also important to examine the percentage of students who reached their targeted goals as well as the percentage of students whose attained ROI reached the benchmark goal for their grade level. Figure 8 shows the percentage of students whose attained ROI was equal to or greater than the student's targeted ROI as well as the percentage of students whose attained ROI was equal to or greater than the benchmark ROI established for their specific grade level. As evident from the figure, the greatest percentage of students meeting these goals was present in kindergarten. Across grades, a higher percentage of students met or exceeded the benchmark ROI for their grade level than met their targeted ROI. An increasing trend of percentage of students meeting their targeted and benchmark ROI for grade was evident across grade level.

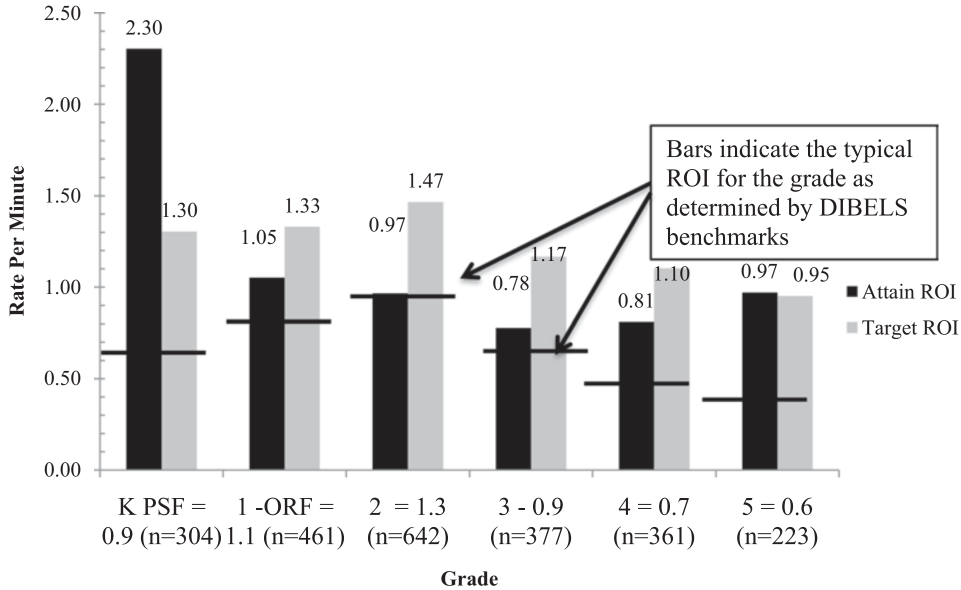


Fig. 7 Attained and target rate of improvement (ROI) across schools and grades, according to Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Data for grades 2, 3, 4, and 5 are based on ORF. (*PSF* phoneme segmentation fluency, *ORF* oral reading fluency)

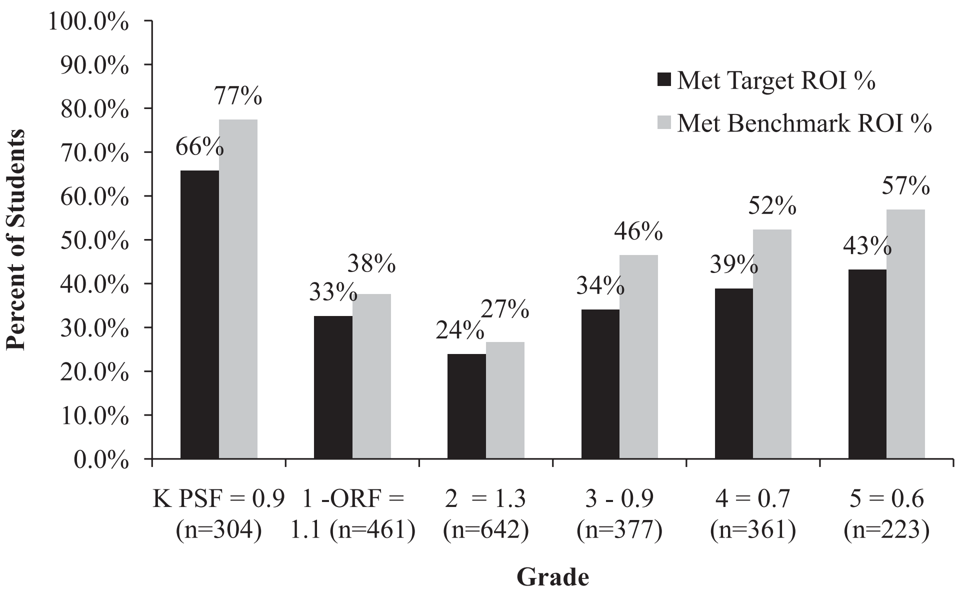


Fig. 8 Percentage of students meeting target and benchmark rate of improvement (ROI). Data for grades 2, 3, 4, and 5 are based on ORF. (*PSF* phoneme segmentation fluency, *ORF* oral reading fluency)

Special Education Referral Patterns

Data were provided from a limited number of schools/school districts regarding special education eligibility decisions. Of the districts solicited, data were obtained from five districts, with one district providing data from ten schools. The data were examined for those students referred for SLD determination, analyzing the number and percentage of students found eligible against those found not eligible.

Examination of these data, although limited in number, showed that between <1 and 2.5% of students across these schools were evaluated for special education as potential students with SLD. Between 33 and 61% of students evaluated for SLD were found to be eligible. Parent initiated referrals varied from approximately 33 to 67% of evaluations. It was important to note that the actual number of students evaluated in all schools for potential eligibility in other categories beside SLD were far greater in all schools.

Given the limited number of schools reporting eligibility data, any conclusions regarding the impact of RTII on referral patterns was very premature at the time of the analysis. However, the richness of the data set provided by one district with ten schools offered an opportunity to do a more in-depth discrepancy analysis between eligible and noneligible students evaluated for SLD evaluation. Across the 14 schools, a total of 52 students were referred and evaluated for an SLD determination. The attained end-of-year universal screening score on ORF for each student was divided by the expected end-of-year benchmark score for that respective grade. In other words, if a first-grade student scored 20 words correct per minute on an ORF assessment during the spring benchmark assessment, the expected benchmark score of 40 was divided by their score, which would result in a discrepancy score of 2.0. A student who scored 1.0 would have no discrepancy, those scoring below 1.0 would have ended up better than the end-of-year benchmark.

A statistically significant (t (45) = 4.27, $p < 0.001$) difference was found between those students found eligible for SLD and those not found eligible. The average discrepancy of those found eligible was 2.42, compared to scores of 1.51 discrepant of those found not eligible, indicating that students who were found eligible were on average 2.5 times discrepant from benchmark compared to those not eligible who were 1.5 times discrepant. Important to note in the analysis is that the standard deviation was almost 2.5 times larger for those found eligible (SD = 1.06) than those found not eligible (SD = 0.44). The difference in variation between groups suggested that there were quite a few students who fell outside the pattern described above for those found eligible, but little variation for those found not eligible. The interpretation of these data suggested that the discrepancy analysis alone could be successful in identifying those determined *to not be eligible* for special education, but the discrepancy analysis alone would not be as effective for identification of those found eligible.

Translating the findings into the decision-making process, teams clearly need to use multiple data sources when making decisions about eligibility, but that the use of a discrepancy analysis might be an initial screen for ruling out special education eligibility for SLD. Those students who are not within −1 SD of the score for those found eligible (discrepancy score of 1.4) would be viewed as unlikely to meet requirements through a comprehensive evaluation for SLD. Those who are at least −1 SD (>1.4 discrepant) would be assessed in greater depth to determine eligibility for special education. Certainly, this analysis requires far more cases to confirm the value of the discrepancy analysis, given that the analysis was based on only 52 cases across five schools, with many cases coming from a single larger district. Future analysis may want to examine these cases in more detail to fully understand the decision-making process in view of the discrepancy analysis.

Implications for Practice and Conclusions

When schools implement an RTI model, it is critically important for the systematic evaluation of the impact of the model on student achievement. Multiple perspectives of the model need to be considered examining changes in identified risk level, shifts in the numbers of students requiring more intensive interventions, responses to interventions among those students receiving supplemental intervention, and the overall effect on the number of students for whom referrals for special education are needed. As discussed by Shapiro and Clemens (2009), one would expect changes in each of these indicators with a collective examination reflecting the degree to which the RTI model is having the desired effect.

The example of the analysis of the implementation of RTI across schools in Pennsylvania provided a series of general findings. The strongest impact of the model was found to be at the earliest grades. Although improvements were evident at every grade, there was much less movement and change in student risk status from grades 2 through 5. Second, most of the change in risk status was found from the fall to winter benchmark periods, than from winter to spring. Third, examination of student progress monitoring of those receiving tier 2 or tier 3 intervention found that students in the youngest grades (kindergarten and first grade) showed a reduction of the gap between themselves and typical performing peers compared to those in higher grades, again reinforcing the finding that the greatest improvements in student achievement occur within the earliest grades. Finally, an analysis of a small subset of schools examining the differences in performance between those found eligible and those not found eligible for special education as a student with SLD.

The example analysis described here would be viewed as a preliminary outcome, but there are potential implications for practice and research as outlined in Table 3. There are also many areas for additional research. First, replicating the findings across a much larger sample of schools, examining outcomes by school demographics, and examining additional indicators of school achievement would be needed to reach definitive conclusions about the impact of the RTI model. At the same time, the type of measures and analysis offers schools, school districts, and state educational agencies the opportunity to conduct strong program evaluation of the effectiveness of RTI. Second, expansion of the evaluation to include examination of the fidelity of the processes of implementation (i.e., was the model implemented as it was proposed to be implemented), examination of the impact of the model on student outcomes on annual state assessments, examination of the acceptability of the model by teachers, parents,

Table 3 Summary table of potential implications for practice

Area	Potential implications for practice based on the current evaluation
Assessment	Oral reading fluency (ORF) measures accurately identified students who were not likely to meet expected levels of reading performance
	ORF data underestimated the levels of proficient reading achievement for grades 3 through 5
	No single measure, such as ORF or 4Sight should be used in making predictions of reading performance for students
	School teams should use multiple data sources to make specific learning disability (SLD) identification decisions
	Discrepancy analysis might be an initial screen for ruling out special education eligibility for SLD. Those students who are not within −1 SD of the score for those found eligible would be viewed as unlikely to meet requirements through a comprehensive evaluation for SLD. Those who are at least −1 SD would be assessed in greater depth to determine eligibility for special education
Intervention	The frequent movement across tiers in early grades reinforced the importance of early intervention and the need to attend to improving literacy skills at the youngest grades
	Student growth tended to stabilize over grades and time once students reached the middle of the school year
Evaluation	The four basic questions focused on risk level described in this evaluation can be used to evaluate the impact of RTI and are an excellent starting point in developing a system-wide evaluation process for RTI

RTI response to intervention

and school administrators, and impact on school policy would further add to a complete evaluation of the impact of an RTI model.

Research regarding methods to evaluate the effectiveness of RTI models is just beginning and these data provide preliminary support. However, using the four basic questions focused on risk level described in this evaluation can serve as an excellent initial starting point in developing a system-wide evaluation of the impact of RTI.

References

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision-making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23,* 362–380.

Bollman, K. A., Silberglitt, B., & Gibbons, K. A. (2007). The St. Croix River Educational District model: Incorporating systems-level organization and a multi-tiered problem-solving process for intervention delivery. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention*: *The science and practice of assessment and intervention* (pp. 319–330). New York: Springer.

Callender, W. A. (2007). The Idaho results-based model: Implementing response to intervention statewide. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention*: *The science and practice of assessment and intervention* (pp. 331–342). New York: Springer.

Clemens, N. H., Shapiro, E. S., Hilt-Panahon, A., & Gischlar, K. L. (2011). Project MP3: Monitoring progress of Pennsylvania pupils: Student achievement outcomes. In E. S. Shapiro, N. Zigmond, T. Wallace, & D. Marston (Eds.), *Models of response-to-intervention implementation*: *Tools*, *outcomes, and implications* (pp. 77–98). New York: Guilford Press.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37,* 293–306.

Gettinger, M., & Stoiber, K. (2008). Applying a response-to-intervention model for early literacy development in low-income children. *Topics in Early Childhood Special Education, 27,* 198–213.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene: Institute for the Development of Educational Achievement. http://dibels.uoregon.edu/.

Hartman, W. T., & Fay, T. A. (1996). *Cost-effectiveness of instructional support teams in Pennsylvania*. *Policy Paper Number 9 of the Center for Special Education Finance*. Palo Alto: American Institutes for Research.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision-making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research and Practice, 18,* 187–200.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71,* 445–463.

Peterson, D. W., Prasse, D. P., Shinn, M. R., & Swerdlik, M. E. (2007). The Illinois flexible service delivery model: A problem-solving model initiative. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention*: *The science and practice of assessment and intervention* (pp. 300–318). New York: Springer.

Shapiro, E. S., & Clemens, N. (2009). A conceptual model for evaluating system effects of RTI. *Assessment for Effective Intervention, 35,* 3–16.

Shapiro, E. S., Solari, E., & Petscher, J. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences, 18,* 316–328.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klingner, & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 11–27). Baltimore: Paul H Brookes Publishing.

Data-Based Decision-Making, the Problem-Solving Model, and Response to Intervention in the Minneapolis Public Schools

Doug Marston, Matthew Lau, Paul Muyskens and Jennifer Wilson

The Minneapolis Public Schools (MPS) has a long history of using data-based decision-making, the problem-solving model (PSM), and response to intervention (RTI). This commitment can be traced to the special education department's adoption of the principles outlined in Deno and Mirkin's (1977) *Data-Based Program Modification*. Deno and Mirkin make the case that evaluating educational services with objective student data contributes to better decisions for students with learning problems. The approach has five major steps: (1) problem identification, (2) problem definition, (3) designing intervention plans, (4) implementing the intervention, and (5) problem solution. Each step is designed to address an important question about the student, which if answered with valid data, leads to effective programming for the student.

For the problem identification step, the question to answer is, "Does a problem exist?" At the next step, problem definition, the question "Is the problem important?" focuses on the student's difficulty inside and outside the classroom and provides information on the extent to which the student is discrepant from expectations. Once the student's problem has been defined and its significance determined, the educator must answer, "What is the best solution hypothesis?" at the designing intervention plan step. At the next step, which addresses implementation of the solution, the educator must determine, "Is the solution attempt progressing as planned?" Finally, at the problem solution step, the question, "Is the original problem solved?" will be answered by examining student data and the response to the intervention.

Data-based decision-making in MPS advanced with the implementation of curriculum-based measurement (CBM; Deno 1985). Since the early 1980s, CBM has been used as an indicator for a variety of important educational decisions for students with mild disabilities, including screening, eligibility, program planning, progress monitoring, and program evaluation (Marston and Magnusson 1985). Implementation of CBM in the district has been expanded into general education, has been used with English language learners (ELL), and is a primary source of data for schools implementing RTI (Deno and Marston 2007; Marston 2012).

The Problem-Solving Model in the Minneapolis Public Schools

Two initiatives that contributed to the district adoption of the PSM were the Six-Week Assessment Plan (SWAP; Marston and Magnusson 1988) and the collaborative teaching project (Self et al. 1991). The SWAP encouraged classroom teachers to implement a general education intervention before making a referral to special

D. Marston (✉) · M. Lau · P. Muyskens
Minneapolis Public Schools, Minneapolis, USA
e-mail: doug.marston@mpls.k12.mn.us

J. Wilson
University of Minnesota, Minneapolis, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_38

education. During this 6-week period, teachers used CBM procedures to measure student growth as a function of the pre-referral strategy that was used with the student. “The purpose of the SWAP is to collect data to determine if any specific interventions in a student’s educational program allows him or her to make appropriate progress in a regular education curriculum” (Marston and Magnusson 1988, p. 153).

These tenets of the PSM became the central focus of an Office for Special Education (OSEP)-funded Model Demonstration Project at Hiawatha Elementary School. At the school, staff implemented a data-based decision-making model where the regular education, Title I, and special education teachers teamed to frequently review the progress of students on CBM procedures and then provided instruction based on student need.

The PSM became a reality in the early 1990 when the state of Minnesota moved to a special education eligibility determination process requiring the administration of a formal measure of intellectual functioning for students with academic needs. At that time, district personnel in the special education department became concerned with the use of intelligence tests as part of the criteria in determining students as “learning disabled” or “mild mentally impaired” especially because the majority of students in MPS were students of color and a large number of the students were ELL. This concern led to MPS seeking a waiver from state rules and gaining approval from the State Board of Education in 1993 to use PSM as an alternative approach to special education eligibility. In this model, MPS followed a variation of the five steps outlined by Deno and Mirkin (1997): (1) definition of the problem, (2) selection and implementation of an intervention, (3) monitoring student progress and response to intervention, and (4) cycle through this sequence if the student is not making adequate growth. The RTI steps are implemented at each of the three stages of the PSM.

Stages of MPS Problem-Solving Model

Stage 1 of the PSM is labeled, “classroom interventions.” At this stage, classroom teachers are asked to define the student’s difficulties, provide baseline data, specify an intervention, and document the results. The classroom intervention worksheet, also known as worksheet 1, is used by staff to document the process. The prompts for the teacher on this worksheet are:

- Review cumulative file/relevant school history
- Talk with staff
- Interview students
- Interview parents
- Document concerns with specificity
- Current levels of performance (baseline data)
- Student strengths
- Relevant health information
- Intervention tried/results
- Start date and follow-up date

Those students who remain discrepant from expectations after approximately 4–6 weeks move on to stage 2 of the PSM. This stage is known as the building team intervention stage where student difficulties are addressed through the building problem-solving team. These teams are typically composed of the regular education teacher, Title I interventionist, school social worker, school psychologist, special education teacher, other specialists as indicated, and a building administrator. Again, the four steps of the PSM are followed in sequence. The major elements of stage 2 are similar to stage 1 with more intensive interventions and more frequent data collection. Staff completes the team intervention worksheet (worksheet 2) which includes the following prompts:

- Primary source of the referral
- Health/additional health-related information, health review, vision/hearing screening
- Parent input, date of contact
- Define/redefine specific behavior concerns
- Current level of performance/baseline data
- Specific goal for intervention

- Intervention plan including outcome goals, results, start date, staff responsible, and follow-up date
- Decision/outcome of the team interventions

The responsibility of the PSM team is to ensure that high-quality interventions are implemented with the student and that the data be reviewed approximately 6–8 weeks after the intervention has been initiated.

While stages 1 and 2 involve cycling through a series of general education interventions and evaluating RTI through progress monitoring, stage 3 marks the beginning of formal due process and special education evaluation. RTI data continues to be collected at this stage and district personnel initiate the processes of completing "Notice of Evaluation/Reevaluation," evaluation plan, evaluation report, and determination of eligibility. These steps are included in the special education evaluation worksheet, or worksheet 3.

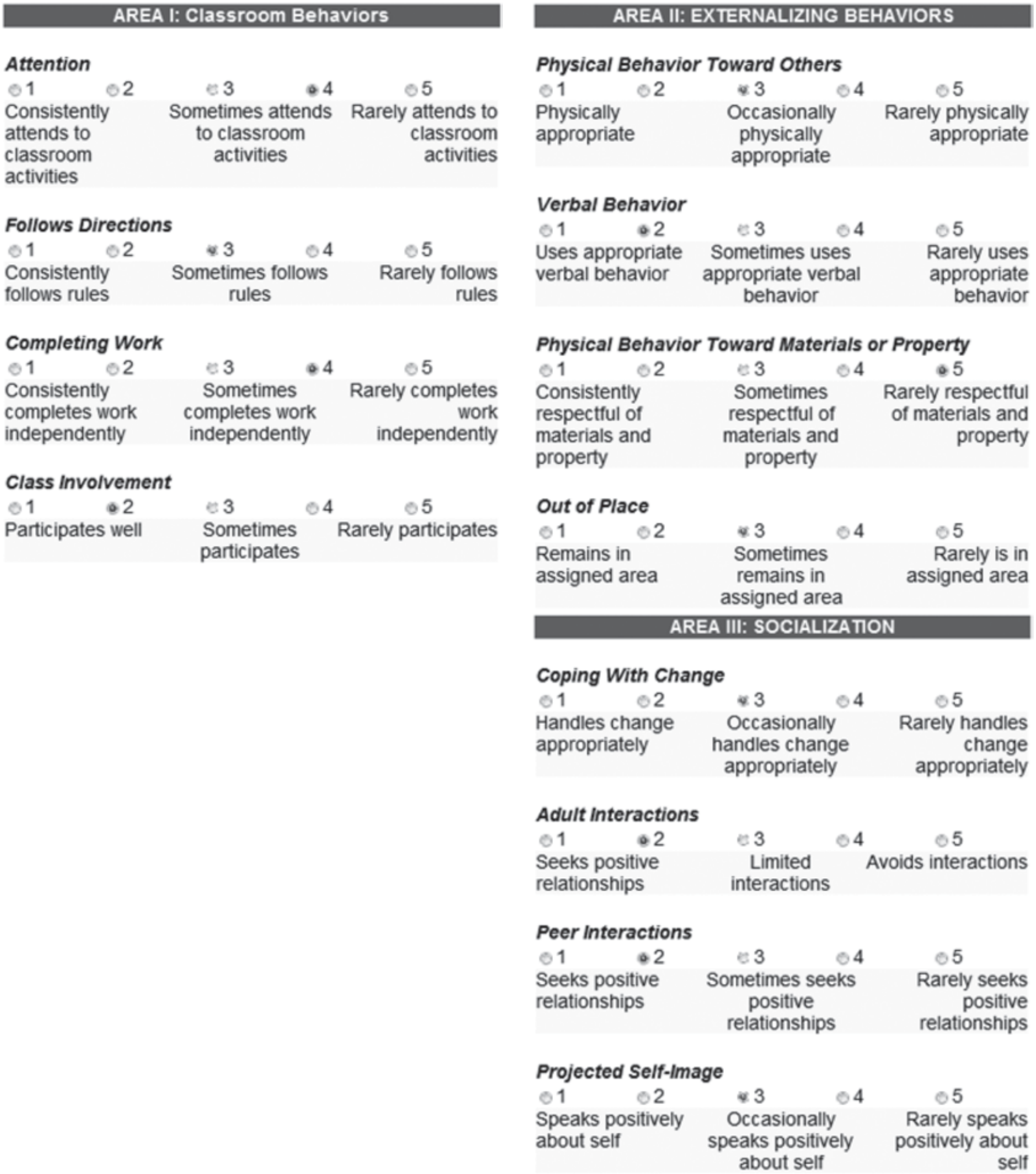
Office for Civil Rights Voluntary Compliance Agreement

Beginning in the 1998–1999 school year, MPS agreed to participate in a voluntary compliance agreement with the United States Department of Education's, Office for Civil Rights (OCR). In this agreement, OCR requested that MPS follow a plan that included: improving screening of all students in academics and behavior, providing a wide range of interventions for students performing poorly in these areas, using a multidisciplinary team at the school to identify interventions, monitoring the progress of these students, and minimizing bias in special education placement by utilizing these concepts. Implementation of this agreement resulted in activities which bolstered the district implementation of PSM.

One of the provisions included in the OCR agreement was universal screening in grades K–8 to identify children who were having difficulties in reading, mathematics, and behavior. Academic screening centered on CBM and use of the (Northwest Educational Association 2003) assessments. In the area of behavior, district staff developed the behavior screening checklist (Muyskens et al. 2007), shown in Fig. 1. The behavior screener is a 12-item checklist in which teachers rate the extent to which students exhibit behavior issues across three domains: classroom behaviors, externalizing issues, and socialization. Teachers rate the student on a scale of 1–5 with higher scorers signifying significant difficulties. The 12 items are attention, follows directions, completing work, class involvement, physical behavior toward others, verbal behavior, physical behavior toward materials or property, out of place, coping with change, adult interactions, peer interactions, and projected self-image.

All students are screened in the fall, with follow-up data collected for identified students and new students in the winter and spring. Students who are below the district-established cutoff scores are considered to be at risk. These students enter the first stage of the PSM. While screening is now commonly accepted as part of any RTI protocol and is an important part of the PSM, at the time it was a new feature of the academic intervention process. The inclusion of all students, teachers, and principals in the process helped facilitate the integration of the PSM into general education.

The district's commitment to data-based decision-making and the PSM led to the creation of a district-wide online data warehouse in 1999, the OCR website. Types of data available to staff include general demographic information and school history; academic information including district and state assessment results and CBM scores; and behavioral information such as behavior screener scores, attendance, and suspension data. These data follow the student from school to school and include a historical record that extends to the student's initial enrollment in the district. Fig 2 is an example of the *summary organizer* for a classroom in which the academic and social/behavioral data are displayed for each student. Based on an analysis by district evaluation staff, cutoff scores identifying students at risk in reading, math, and behavior are applied to individual student data. A three-color coding system (red, yellow, and green) is used to help staff members identify at-risk students visually so that further interventions can be implemented in the PSM. Green signifies that students have met or exceeded standards, yellow represents partial mastery of standards, and red signals the



AREA I: Classroom Behaviors

Attention

1	2	3	4	5
Consistently attends to classroom activities		Sometimes attends to classroom activities		Rarely attends to classroom activities

Follows Directions

1	2	3	4	5
Consistently follows rules		Sometimes follows rules		Rarely follows rules

Completing Work

1	2	3	4	5
Consistently completes work independently		Sometimes completes work independently		Rarely completes work independently

Class Involvement

1	2	3	4	5
Participates well		Sometimes participates		Rarely participates

AREA II: EXTERNALIZING BEHAVIORS

Physical Behavior Toward Others

1	2	3	4	5
Physically appropriate		Occasionally physically appropriate		Rarely physically appropriate

Verbal Behavior

1	2	3	4	5
Uses appropriate verbal behavior		Sometimes uses appropriate verbal behavior		Rarely uses appropriate behavior

Physical Behavior Toward Materials or Property

1	2	3	4	5
Consistently respectful of materials and property		Sometimes respectful of materials and property		Rarely respectful of materials and property

Out of Place

1	2	3	4	5
Remains in assigned area		Sometimes remains in assigned area		Rarely is in assigned area

AREA III: SOCIALIZATION

Coping With Change

1	2	3	4	5
Handles change appropriately		Occasionally handles change appropriately		Rarely handles change appropriately

Adult Interactions

1	2	3	4	5
Seeks positive relationships		Limited interactions		Avoids interactions

Peer Interactions

1	2	3	4	5
Seeks positive relationships		Sometimes seeks positive relationships		Rarely seeks positive relationships

Projected Self-Image

1	2	3	4	5
Speaks positively about self		Occasionally speaks positively about self		Rarely speaks positively about self

Fig. 1 Behavior screener

GR	Gender	Ethnicity Code	ELL	SpEd	CBM Rdg. F11	CBM Rdg. W12	CBM Rdg. S12	MAP-Rdg. F11	MAP-Rdg. W12	MAP-Rdg. S12	Reading Tier	CBM-Math F11	CBM-Math W12	CBM-Math S12	MAP-Math F11	MAP Math W12	MAP Math S12	Math Tier	Behavior Screener F11
3	M	4	Y		24	47	88	167	177	182	3	35	36	35	185	197	202	2.	12
3	M	3	Y		127	148	153	198	207	209	1	38	43	43	202	208	220	1.	30
3	M	5		SLI	82	91	102	203	208	217	1	38	45	41	200	216	213	1.	28
3	F	2			54	77	89	184	184	188	2	32	34	37	184	192	190	2.	18
3	F	3			109	139	141	203	205	203	1	35	36	44	196	209	211	1.	26
3	M	4	Y		79	86	110	186	195	196	1	38	41	39	202	205	211	1.	14
3	M	4	Y		77	106	112	182	196	197	1	27	36	41	191	199	197	2.	23
3	M	2			89	108	111	206	210	209	1	45	45	45	212	208	217	1.	28
3	M	2			72	84	103	177	194	178	3	22	35	37	180	192	187	3.	12
3	F	3	Y		55	75	101	178	193	195	1	31	35	34	188	192	203	2.	26
3	F	2			81	100	124	204	215	196	1	35	41	40	202	206	212	1.	27
3	M	3	Y		51	73	95	176	184	181	3	33	35	39	181	193	196	2.	18
3	F	4			78	100	111	190	197	205	1	36	39	30	199	199	202	1.	14
3	M	2			31	52	68	163	152	165	3	23	25	27	159	163	173	3.	29
3	F	4	Y	SNAP	26	41	70	168	153	201	2	23	29	25	170	171	187	3.	12
3	M	2			50	113	111	168	155	195	1			32	171	166	188	3.	12
3	M	2			29	67	75	190	192	198	2	34	39	40	199	200	209	1.	24
3	M	3			86	101	112	189	194	186	2	38	37	37	188	193	195	2.	22

Fig. 2 A summary organizer that displays academic and social/behavioral data for students in a classroom

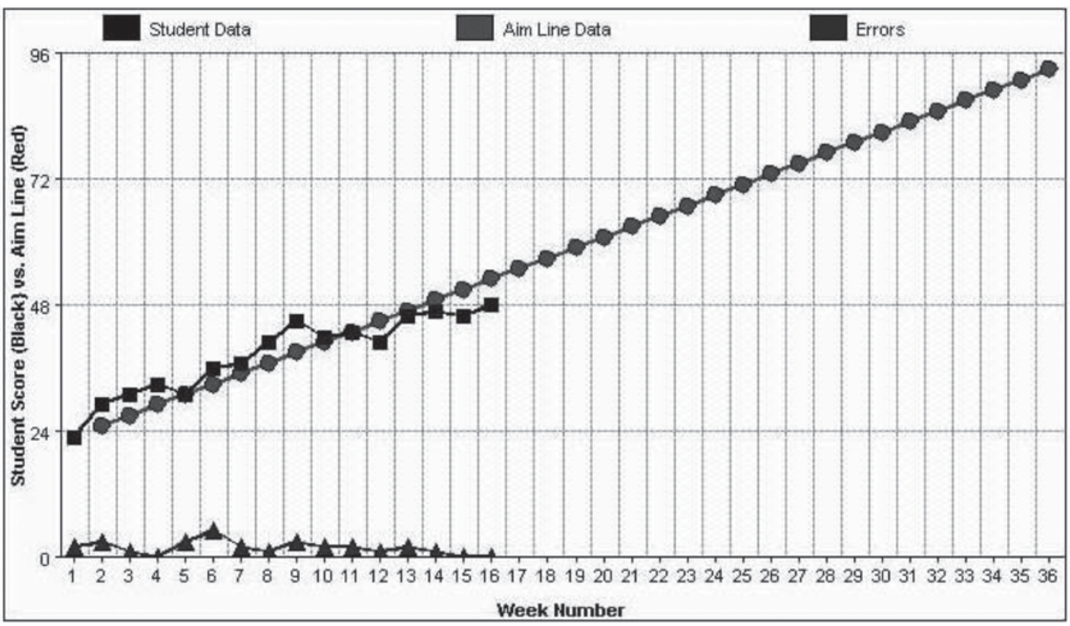


Fig. 3 Example of a web-based individual student progress monitoring graph

student has not met standards and is in need of intensive intervention. In addition to viewing screening data, teachers are able to input progress-monitoring data as shown in the example of a progress-monitoring chart for words read correctly in Fig. 3. Teachers can also document their interventions and describe specific details of how service is delivered.

What We Have Learned About the PSM

Having implemented the PSM since the mid-1990s, an understanding of how theory and research on this model translates into everyday practices in the schools has been gained. In the remainder of this section, the authors review what they have learned about student outcomes, roles and responsibilities, integration with general education, and the use of data.

Student Outcomes

One of the concerns expressed when the district initially implemented problem-solving was the possibility the model would lead to an increase in the number of students with mild disabilities. It was found that using the PSM did not have a significant effect on the prevalence of students with mild disabilities in MPS. The initial waiver from the Minnesota Department of Education allowed the authors to use the PSM for the special education categories of learning disabilities and mild mental impairment. As reported in Marston et al. (2003), the percentage of students identified with a learning disability or mild mental impairment was consistently around 7 % of the student population over several years. The data show that implementation of PSM did not affect this percentage.

Another question that is often asked about PSM is whether different types of students are identified with this model versus the traditional discrepancy model. The data reported in Marston (2002) showed similarities in achievement levels of students found eligible with PSM and those labeled using the intelligence quotient (IQ)–achievement difference score.

The quality of the pre-referral interventions used in the PSM was addressed in a study conducted by Reschly and Starkweather (1997). These researchers concluded the interventions used by teachers using the PSM were typically of higher quality than those provided under the traditional eligibility approach. These researchers also reported that students in the PSM were provided service in special education at an earlier age than students identified with traditional criteria.

Roles and Responsibilities of Staff

The authors have identified key staff required for successful implementation of the PSM. Stakeholders include, but are not limited to, district and school-level administrators, general education teachers, special education teachers, and school psychologists. District and school-level administrators, including principals, are major players in sustained quality implementation of the PSM. These leaders provide vital support by establishing a common vision, incorporating data-based decision-making into the current school system, providing tangible (e.g., budget allocation for materials, intervention programs, staff time) support for the process, and performing essential administrative functions to ensure staff members remain on course (Lau et al. 2006).

Response to Intervention in the Minneapolis Public Schools

RTI is depicted as multilayered system of instructional opportunities that are aligned to student instructional needs (Fuchs and Fuchs 2006). The essential components of successful RTI implementation include: valid and reliable assessments that measure student level of performance and growth; evidence-based instruction and intervention methods; and a framework for implementing screening, targeted interventions, progress monitoring, and data-based decision-making with fidelity (Vaughn et al. 2007). The district's history with data-based decision-making, the PSM, and implementation of the OCR agreement provided the foundation for RTI implementation. MPS has been acknowledged as an early implementer of RTI because of its PSM (Stepanek and Peixotta 2009) and has shown effectiveness (Burns et al.

2005; Burns and Ysseldyke 2005). In 2007, in collaboration with the University of Minnesota, MPS was funded by the US OSEP to build upon these prior experiences to demonstrate RTI at three elementary schools (Marston et al. 2011b; Wallace et al. 2011). The authors re-branded PSM and incorporated the popular terminology of RTI. Stage 1: Classroom interventions of PSM became tier 2 at the demonstration schools. Stage 2: Problem-solving team interventions became tier 3. Stage 3: Special education evaluation was incorporated into tier 3.

An additional feature added to the model was grade-level data teams, a team that convenes earlier than the building PSM. This team views the data of all students on a regular basis, once every 6 weeks. At these meetings, an in-depth review of tier 1 core instruction and its impact upon the student is conducted as well a discussion of tier 2 and 3 interventions. An example of the tiers of instruction and intervention at kindergarten at the demonstration school is presented in Table 1. In time, school lead staff noted the grade-level data team and professional learning communities (PLCs) had a common purpose. This alignment provided the building administration the opportunity to merge both school meetings.

Screening and Progress Monitoring

Essential to the adoption of the PSM is the use of data from several sources and at various points in time during screening and progress monitoring. The RTI demonstration sites currently use a variety of assessments for screening purposes including NWEA measures of academic progress (MAP) and CBMs. As described earlier and shown in Fig. 2, the screening data are organized in an easy-to-analyze format for determining those individual students who extra help. If the data show a student may not be on track to be proficient on the state high-stakes assessment, additional diagnostic assessments are conducted. Drop-line charts, as shown in Fig. 4, also provide a visual representation of CBM classroom data that allows staff to determine student growth toward proficiency on state accountability measures.

District-developed CBM procedures are used for progress monitoring. All staff have access to a document (MPS, n.d.) that includes an introduction to the use of CBM, directions for administration and scoring of measures, and passages and probes to be used for oral reading, reading expression, comprehension questions, and story retells. District norms and data relating these measures to state assessments are also included in this manual (Tindal and Marston 1996; Muyskens and Marston 2002; Marston et al. 2003). In addition, district staff developed early literacy measures that assess letter–sound fluency, onset phoneme identification, and phoneme segmentation (Marston et al. 2007). retain as 2007

Data Meetings

An important feature of an RTI implementation, and not present on traditional PSM implementation, was the use of grade-level data teams. The authors' experience indicates data meetings are important to change instruction and impact student achievement. The staff developed a grade-level data team fidelity checklist (Marston and Lau 2008) to serve as a guideline for each meeting:

1. Agenda and expectations reviewed
2. Meeting is at least 90 min long
3. Students in "red" and "yellow" zones are reviewed
4. Teachers describe instructional methods and materials for targeted groups
5. Student graphs were complete
6. Student graphs were reviewed
7. Benchmark data reviewed
8. Other assessment data reviewed
9. Team input
10. Student decision related to criterion
11. Target intervention is adjusted as needed

In addition, the checklist was used at the RTI demonstration schools to judge the integrity of the data review meetings.

Table 1 Example of interventions provided in one Minneapolis Public School for kindergarten students

Focus group teacher	Intervention plan	Tier 1	Tier 2	Tier 3
Special education	Meet in a small group of four students daily for 45 min/day. Use movement learning, quick run through, and highlighting to review letter sounds and sight words. Read texts 2–3 times per week focusing on concept of word and beginning sounds. On Mondays preview text, content, and vocabulary for that week's core curriculum text			xx
ELL teacher	Meet in a small group of 3–4 students daily for 20 min/day. Support language development through word walls, visual supports, pair speaking activities, and vocabulary games. On Mondays preview text, content, and vocabulary for that week's core curriculum text	xx	xx	xx
Associate educator	Meet in a small group of 3–4 students daily for 20 min/day. Use direct instruction—*SRA Reading Mastery*. Include rhyming and sight word instruction			xx
Kindergarten classroom teacher	Meet in a small group of four students daily for 45 min/day. Develop letter sounds using Success for All curriculum and videos, and sound tubs, and begin to build words. Use level A (instructional) text to develop concepts of print. On Mondays, preview text, content, and vocabulary for next week's core curriculum text			xx
Kindergarten classroom teacher and associate educator	Meet in a group of 19 students daily for 45 min/day. Use K-PALS program (whole group and partners) to develop phonemic awareness and phonics. Pairs pulled for additional phonemic awareness or phonics interventions (from MN Reading Corps) administered as per student need. Add in books with high-frequency words, word family, etc. On Mondays preview text, content, and vocabulary for that week's core curriculum text		xx	
Math specialist	Small group of six students meet daily for 45 min/day. Literacy: Administer 5-min letter–sound review. Read leveled text related to math concepts. Math: "Flexible 5"- structured understanding of 5 s and 10 s. Number ID. Students do writing reflection of learning		xx	
Associate educator	A group of 11 students meets daily for 45 min/day. The students meet in small groups to read and respond to text. In addition, the students play word study games that align with *Words Their Way* developmental spelling stages. Additional focus on handwriting (letter recognition and handwriting video, handwriting sheets and play dough), and structuring sentences	xx	xx	
Kindergarten classroom teacher	Meet in a group of 19 students for 45 min/day. Extend student literacy skills and comprehension strategies through partner and independent reading, small group instruction (4–6 students) with instructional level text, and written response to reading. Small groups will focus on handwriting, advanced decoding strategies, self-monitoring. More open writing	xx		

ELL English language learner, *K-PALS* kindergarten peer-assisted learning strategies

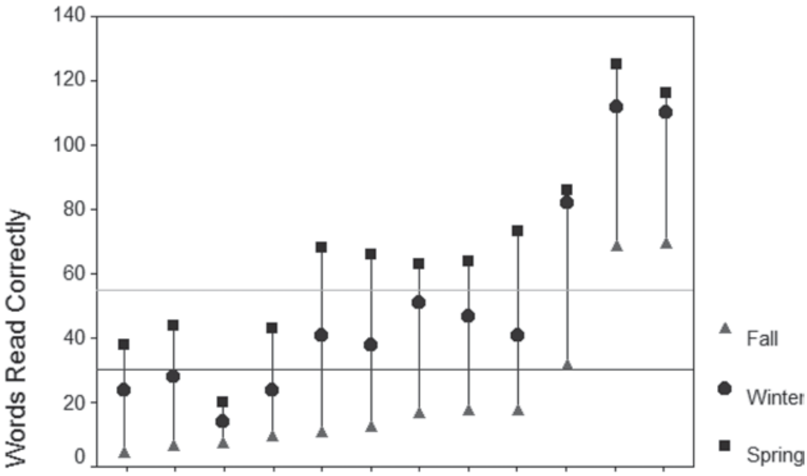


Fig. 4 Example of a classroom CBM graph showing fall-to-winter-to-spring growth for each student

What Have We Learned from the RTI Demonstration Project

This section describes four areas of research originally investigated at the OSEP demonstration schools and report on a 2-year follow-up study of model implementation, fidelity of data teams, staff attitudes toward RTI, and student achievement. All four areas address the issue of sustainability of implementation efforts.

Implementation of the Model

The *RTI Readiness Checklist* (Pennsylvania Department of Education 2007) was used for measuring the degree of implementation at the demonstration sites. This assessment tool comprises ten important domains associated with successful implementation: high-quality curriculum (HQ), universal screening (US), shared ownership (SO), data-based decision-making (DB), tiered interventions (TI), parent engagement (PE), behavior (B), eligibility determination (ED), leadership (L), and professional development (PD). Within each domain, there are two to six items to be evaluated. Project staff, with input from the building principal and school lead staff, rated the extent to which the school was implementing these essential components of RTI. A rubric score of 1–3 is assigned to each of the items in each domain and a total domain score was determined. The percentage of total possible points, which is viewed as the degree of implementation, was then calculated for each domain. The degree of RTI implementation at the demonstration site ranged from 42 to 100 % on the ten domains with an average of 84 %.

The authors did a 2-year follow-up at one of their demonstration sites (school A), which was considered by project staff as a high implementer for RTI. In the last year of the demonstration project (2009), the school was assigned 100 % of possible points on the domains US, SO, DB, and eligibility which is considered full implementation. The school also scored high on TI (93 %) and L (95 %). The lowest percentage of implementation points was found for PE (40 %) and B (60 %). The authors asked the question, was the school able to sustain RTI implementation without the resources of the federal grant? To answer this question they used the RTI checklist to measure implementation 2 years after the grant had expired. As shown in Fig. 5, average ratings, although slightly lower with ratings ranging from 33 to 100 %, (average = 76 %) remained high; evidence the school was able to sustain their implementation process.

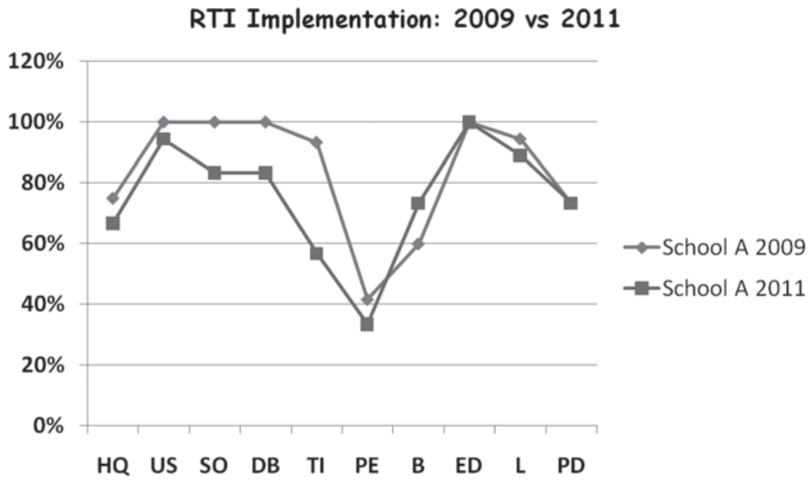


Fig. 5 Percentage of possible implementation points as measured on the RTI implementation checklist in 2009 and 2011. (Note *HQ* high-quality curriculum, *US* universal screening, *SO* shared ownership, *DB* data-based decision-making, *TI* tiered interventions, *PE* parent engagement, *B* behavior, *ED* eligibility determination, *L* leadership, and *PD* professional development)

Fidelity of Data Meetings

Project staff reviewed the fidelity of the grade-level data teams with the data review meeting integrity checklist (Marston and Lau 2008), described earlier in this chapter. Observers use a five-point scale to rate 11 elements: agenda and meeting expectations (E1), length of meeting (E2), review of students in "red" and "yellow" zones (E3), description of interventions (E4), completeness of progress-monitoring graphs (E5), reviewing progress-monitoring data (E6), reviewing benchmark data (E7), reviewing other data sources (E8), team input (E9), criteria used in decision-making (E10), and intervention changes based on student data (E11). Average domain ratings during the project ranged from 4.2 to 5.0 with a mean average of 4.6, indicating a high degree of fidelity in the data meetings (Wallace et al. 2011). Again, sustainability reviewing data team fidelity 2 years after the grant had expired was examined. Integrity was high again with average ratings ranging from 3.3 to 5.0 with an average of 4.4. These data are plotted in Fig. 6.

Staff attitudes toward RTI were measured at the end of the demonstration project by teacher survey and reported by Wallace et al. (2011). The questionnaire asked teachers to use a four-point scale to rate their agreement with statements regarding the major components of RTI: screening, target interventions, progress monitoring, and data review meetings. The following statements for each implementation area were listed for teachers on the survey.

Screening

1. Improves instructional programming
2. CBM and benchmarks improve decision-making
3. OCR website is a useful tool

Target Interventions

1. Benefits low-performing students
2. Benefits average-performing students
3. Benefits high-performing students

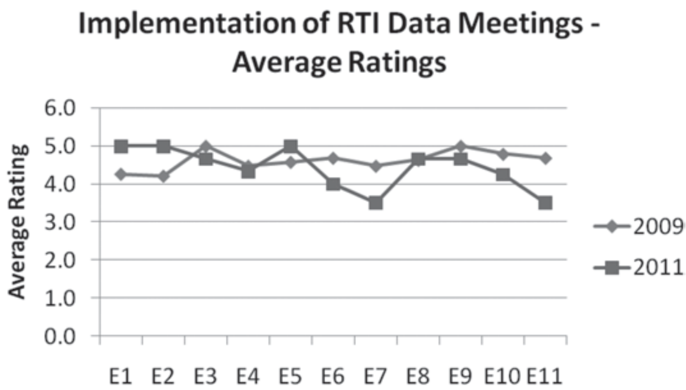


Fig. 6 Average observer ratings for data meeting fidelity in 2009 and 2011. (Note *E1* agenda and meeting expectations, *E2* length of meeting, *E3* review of students in *red* and *yellow* zones, *E4* description of interventions, *E5* completeness of progress-monitoring graphs, *E6* reviewing progress-monitoring data, *E7* reviewing benchmark data, *E8* reviewing other data sources, *E9* team input, *E10* criteria used in decision-making, and *E11* intervention changes based on student data)

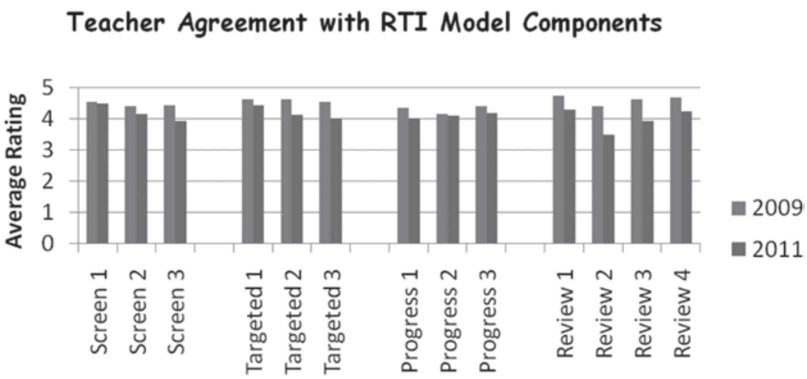


Fig. 7 Teacher attitudes toward RTI components in 2009 and 2011

Progress Monitoring

1. Informs instruction
2. Should be used with all students not meeting standards
3. Useful for forming intervention groups

Data Review Meeting

1. Improves instructional planning for students
2. Meetings should be held monthly
3. Meaningful discussions about research-based interventions
4. Increases collaboration

In general, teacher ratings were high. During the last year of the demonstration project, all mean ratings at school A were above 4.0 with an average of 4.4. The 2-year follow-up, examining the sustainability shows similar, although slightly lower, ratings with an average of 4.1. The degree to which staff agreed with the basic principles of RTI implementation for both years, as provided in Fig. 7, indicates the school had succeeded in building capacity for maintaining the model.

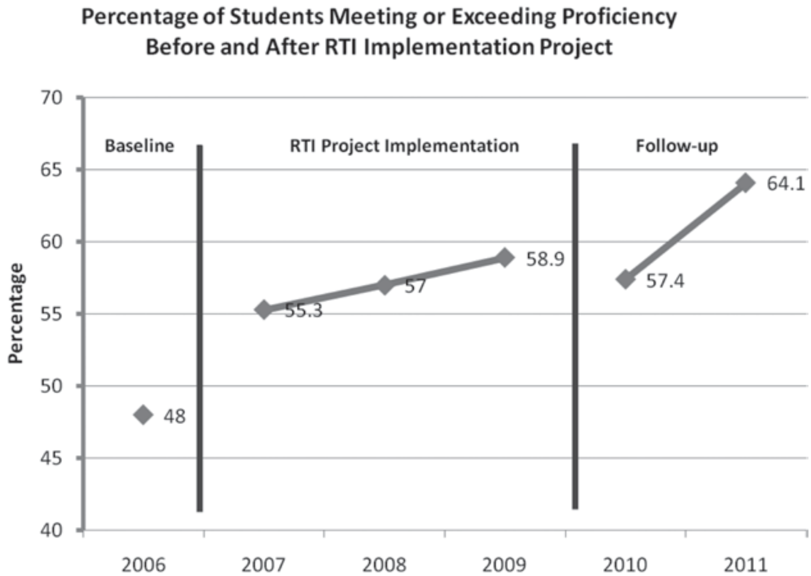


Fig. 8 Student achievement outcomes before, during, and after RTI demonstration project. *RTI* response to intervention

Student Achievement

Student reading achievement was evaluated at the end of the project at school A (Wallace et al. 2011) and 2 years later after completion of the grant. Performance in this area was measured with the state accountability test, the Minnesota Comprehensive Assessment (MCA) test of reading. As shown in Fig. 8 , the percentage of students that had met or exceeded the standard of proficiency before RTI implementation was 48 %. During the 3-year OSEP demonstration project, the proficiency rates improved to 53.3, 57, and 58.9 %. After the project was completed, the proficiency rate dropped slightly to 57.4 % but rebounded to 64.1 % when the 2-year follow-up study was conducted. In a comparison of this school to others in the district, a value-added analysis indicates it performed above predicted achievement levels, when controlling for variables such as special education status, free and reduced lunch status, and race (Heistad 2011).

Implications for Practice

The major elements of RTI implementation have been described as US, evidence-based interventions for students need support, progress monitoring, data-based decision-making teams, and fidelity checks for all RTI components (Elliot and Morrison 2008). This chapter describes how MPS has implemented these components in the PSM and now with RTI implementation. By the end of the RTI demonstration project, the following benefits were identified by school staff:

- "Instructional time engages students more because staff plans have improved and is more focused on what needs to be taught and how.
- Behavior referrals and time off task have decreased.
- Teachers have become more collegial, with more discussion in the staff lounge about instructional issues.
- The staff is excited about learning as adults.
- Special education referrals are lower because the quality of instruction has improved in tier 1 and evidence-based strategies are used in tiers 2 and 3 (Marston et al. 2011a, pp. 243–244)."

Further, the 2-year follow-up study demonstrated that sustainability can be accomplished. However, successful RTI implementation is not limited to the presence of the major RTI elements, implementers need to consider other important factors that have implications for practice. In conducting the follow-up study of RTI implementation, the

Table 2 Summary table of implication for practice for Response to Intervention (RTI)

Area	Implications for practice
Professional development	Focus professional development on evidence-based instruction, assessment, and progress monitoring
	Teachers should meet on a frequent basis to study research, talk about instruction, look at student data, and share their instructional experiences as embedded development
Integration with general education	General education should own academic difficulties and instructional interventions
	RTI models should emphasize quality core instruction and sharing instructional ideas at data team meetings
Use of data	It is important to get the data into the hands of those who need it, often times via data meetings
	Use existing test scores, background information, and prior intervention data to make decisions
	Use an electronic data warehouse to facilitate use of data
Leadership	Principals should be committed to evidence-based practices, data decision-making, and the belief that every student can learn
	The school principal should participate in team meetings that discuss student interventions and progress
	The principal should model collaboration

authors have noted several important factors that deserve the attention of those who adopt these types of models, including professional development, the integration of general and special education, the use of data, and leadership. These are described below and are included in Table 2.

Professional Development

Providing the professional development necessary for successful implementation can be challenging (Butler 2009). Teachers play an important role in affecting student achievement and must have the knowledge, skills, and ability to implement evidence-based instruction (Taylor et al. 2003). Too often, there is a gap between the research that documents effective practices and what actually occurs in typical classrooms (Cook and Cothren Cook 2011). Beyond that, there is a failure to link selection and implementation of effective teaching strategies with student achievement (Foorman 2007; Foorman and Schatschneider 2003). For general education teachers, the emphasis on effective instruction and academic outcomes for students have highlighted the need to further expand their skills in the area of evidence-based instruction, assessment, and progress monitoring for diverse students. One way to address this need is through the PLC model where a culture of collaboration is created and there is a focus on student data (Dufour 2005). At the demonstration school, teachers met on a frequent basis to study research, talk about instruction, look at student data, and share their instructional experiences in PLCs (Marston et al. 2011b).

Integration with General Education

Another area which needs to be an ongoing and high priority for the successful implementation of the PSM is the ownership of academic difficulties and instructional interventions by general education. In MPS, the PSM originally arose from within special education, which resulted in the widely held perspective that the PSM was part of the referral process to get a student special education services. While the OCR made it clear that the origin of special education placements is the referral from general education, it is a perspective that needs to be continually reinforced. RTI's emphasis on quality, core instruction, and the sharing of instructional ideas at data team meetings and PLCs helps support that perspective.

The Use of Data

An area which is vital to the implementation of RTI is the importance of making instructional decisions based on the collection of screening, progress monitoring, and formative assessment data. Shepherd (2006) has identified the development of data-based decision-making culture as essential. However, it must be noted that a data-based decision-making model cannot be successful unless the data is in the hands of those who need it, often times via data meetings. All too often, students must wait to fail before interventions are implemented, even though extensive test scores, background information, and prior intervention data is available. Currently the web-based data warehouse and collection system has helped in this area, but work remains to be done. New tests, graphing features, and connecting data to interventions continue to be ongoing needs.

Leadership

From the authors' perspective, a key factor in the successful implementation of RTI is the leadership provided by the school principal. This view is shared by Shepherd (2006) who identified several characteristics of the effective RTI principle including (a) creates school teams that are interdisciplinary, (b) participates in the school teams that meet to discuss student interventions and progress, (c) models collaboration in the data team meetings and RTI implementation, and (d) ensures there is connection between the RTI teams, professional development, and PLCs.

Conclusion

In general, the data showed that RTI implementation efforts were maintained, data review meetings were implemented with fidelity, teacher attitudes stayed positive, and MCA data improved. Qualitative feedback from the principal indicated RTI implementation has continued with fidelity. Some of the positive contributions of RTI noted by the principal included: (a) a structured system remained in place for all students, (b) there was accountability for all students to ensure learning, and (c) combining RTI and PLC meetings created opportunity for collaboration. A major challenge cited by the principal was school budgets provide limited resources for "red zone" interventions for students needing more intensive services. The principals at the RTI sites displayed the characteristics of effective leadership, and have shown a strong commitment to evidence-based practices and data decision-making. Above all, the PSM and RTI initiatives at MPS facilitated a belief system that every student can learn, and that the students benefited.

References

Burns, M. K., & Ysseldyke, J. E. (2005). Comparison of existing response-to-intervention models to identify and answer implementation questions. *California School Psychologist, 10,* 9–20.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381–394.

Butler, L. (2009). A step-by-step guide to response to intervention. *Principal, 89*(1), 46–52.

Cook, B. G., & Cothren Cook, S. (2011). Thinking and communicating clearly about evidence-based practices in special education [White paper]. http://education.uoregon.edu/uploads/1087/Thinking_and_Communicating_Clearly_About_Evidence-based_Practices_in_Special_Education.pdf.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Deno, S. L., & Marston, D. (2007). Curriculum-based measurement of oral reading: An indicator of growth in fluency. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (pp. 179–203). Newark: International Reading Association.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification*. Reston: Council for Exceptional Children.

Dufour, R. (2005). What is a professional learning community? In R. Dufour, R. Eaker, & R. Dufour (Eds.), *On common ground: The power of professional learning communities* (pp. 31–43). Bloomington: Solution Tree.

Elliot, J., & Morrison, D. (2008). *Response to intervention blueprints: District level edition*. Alexandria: National Association of State Directors of Special Education.

Foorman, B. R., & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observation of teaching and learning* (pp. 1–30). Baltimore: Brookes Publishing Co.

Foorman, B. R. (2007). Primary prevention in classroom reading instruction. *Teaching Exceptional Children, 39*(5), 24–30. http://web.ebscohost.com.ezp1.lib.umn.edu/ehost/pdfviewer/pdfviewer?vid=2&hid=13&sid=2aa2f76e-4e1f-4729-a563-486c2907143c%40sessionmgr12.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93–99.

Lau, M. Y., Sieler, J. D., Muyskens, P., Canter, A., VanKeuren, B., & Marston, D. (2006). Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large midwest urban school district. *Psychology in the Schools, 43,* 117–127.

Marston, D. (2012). School-based and district-wide applications. In C. A. Espin, K. L. McMaster, S. Rose, & M. Miura Wayman (Eds.), *A measure of success*: The influence of curriculum-based measurement on education (pp. 59–78). Minneapolis: University of Minnesota Press.

Marston, D., & Lau, M. (2008). *Data review meeting integrity checklist*. Unpublished manuscript. Minneapolis, MN.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52,* 266–276.

Marston, D., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. Graden, J. Zins, & M. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137–172). Washington, DC: National Association of School Psychologists.

Marston, D. (2002). A functional and intervention-based assessment approach to establishing discrepancy for students with learning disabilties. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilties* (pp. 437–447). Mahwah: Erlbaum.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision-making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research and Practice, 18,* 187–200.

Marston, D., Pickart, M., Reschly, A., Muyskens, P., Heistad, D., & Tindal, G. (2007). Early literacy measures for improving student reading achievement: Translating research into practice. *Exceptionality, 15*, 97–118.

Marston, D., Casey, A., & Wallace, T. (2011a). Implementation of RTI in the Minneapolis public schools: Context and content. In E. S. Shapiro, N. Zigmond, T. Wallace, & D. Marston (Eds.), *Models of response-to-intervention implementation: Tools, outcomes, and implications*. New York: Guilford Press.

Marston, D., Wallace, T., Thompson, J., Lau, M., & Muyskens, P. (2011b). Implementation of RTI in a large, urban school district: Process. In E. S. Shapiro, N. Zigmond, T. Wallace, & D. Marston (Eds.), *Models of response-to-intervention implementation: Tools, outcomes, and implications*. New York: Guilford Press.

Minneapolis Public Schools. (n.d.). *Performance assessment of academic skills in the problem-solving model*. Minneapolis, MN: author. September 22, 2006. http://pic.mpls.k12.mn.us/Performance_Assessment_Manual.html.

Muyskens, P., & Marston, D. (2002). *Predicting success on the Minnesota Basic Skills Test in reading using CBM*. Unpublished manuscript, Minneapolis Public Schools.

Muyskens, P., Marston, D., & Reschly, A. L. (2007). The use of response to intervention practices for behavior: An examination of the validity of a screening instrument. *The California School Psychologist, 12,* 31–45.

Northwest Evaluation Association. (2003). *Technical manual for the NWEA Measures of Academic Progress and Achievement Levels Tests*. Portland: Author.

Pennsylvania Department of Education. (2007). *Response to intervention (RTI) readiness and implementation: Self assessment tool*. Harrisburg: Author.

Reschly, D., & Starkweather, A. (1997). *Evaluation of an alternative special education assessment and classification program in the Minneapolis public schools*. Ames: Iowa State University.

Self, H., Benning, A., Marston, D., & Magnusson, D. (1991). Cooperative teaching project: A model for students at risk. *Exceptional Children, 58*, 26–34.

Shepherd, K. G. (2006). Supporting all students: The role of school principals in expanding general education capacity using response-to-intervention teams. *Journal of Special Education Leadership, 19*(2), 30–38.

Stepanek, J., & Peixotto, K. (2009). *Models of response to intervention in the Northwest region states* (Issues & Answers Report, REL 2009-No. 079). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. www.ies.ed.gov/ncee/edlabs.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal*, *104*(1), 3–28. http://login.ezproxy.lib.umn.edu/?url=http://search.proquest.com/docview/62182778?accountid=14586.

Tindal, G., & Marston, D. (1996). Technical adequacy of alternative reading measures as performance assessments. *Exceptional Children, 58*, 26–34.

Vaughn, S., Wanzek, J., Woodruff, A., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klinger, & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 11–27). Baltimore: Brookes.

Wallace T., Marston, D., Ticha, R., Lau, M., & Muyskens, P. (2011). Description of outcomes of the data on individual students, teachers, and integrity of implementation. In E. S. Shapiro, N. Zigmond, T. Wallace, & D. Marston (Eds.), *Models of response-to-intervention implementation: Tools, outcomes, and implications*. New York: Guilford Press.

Implementing Response to Intervention in a Rural Setting

Renee Guy, Amanda Fields and Lynn Edwards

It seems that rural schools are often forgotten in the continued surge to reform and improve education in the USA. Yet, these schools continue to serve a considerable portion of the student population across this country. Approximately 33 % of public K–12 schools nationwide are located in rural school districts and 25 % of all students in public K–12 schools in 2010–2011 attended rural schools (Keaton 2012). Rural schools face unique challenges in serving their student population and focusing on understanding how to best meet the needs of rural schools remains a concern as we look to improve the educational achievement and emotional well-being of all students across the USA.

Characteristics of rural schools across the 50 states generally show that rural schools typically serve a higher proportion of white students, a smaller student population per school, and have a smaller proportion of students qualifying for free and reduced in comparison to urban schools (National Center on Educational Statistics [NCES] 2010). However, student characteristics vary across geographic regions across the nation, and the population of racial/ethnic minorities and English learners served in rural schools is continuing to diversify and increase (Aud et al. 2013). Additionally, approximately 12 % of students in rural areas receive special education services compared to the national average of 8.7 % (Aud et al. 2013).

In terms of academic achievement, rural schools frequently fair better than urban areas but below suburban students on the National Assessment of Educational Progress (Aud et al. 2013). It may also be more difficult for small schools and districts, because the average school in rural areas serve fewer students, to rise above adverse conditions and improve student achievement as well as narrow the achievement gap between poorer and more affluent students (Bickel and Howley 2000; Johnson 2004). Additionally, rural students have a slightly lower high school graduation rate (77 %) in comparison to the national rate (78.2 %) and are less likely to pursue postsecondary education than students in urban settings (Aud et al. 2013; Strange et al. 2012; Provasnik et al. 2007).

Some of the common challenges often discussed as being specific to rural education include staff recruitment and retention, serving students with disabilities, transportation, technology and resource limitations, funding, meeting federal accountability requirements, and college enrollment rates (Barley and Beesley 2007; Monk 2007). Moreover, staff recruitment and retention remains a problem for many reasons in rural education, but often because limited finan-

R. Guy (✉) · A. Fields
Luverne Public Schools, Luverne, USA
e-mail: r.guy@isd2184.ne

L. Edwards
University of Minnesota, Minneapolis, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_39

cial resources make it difficult to pay teachers fairly, provide adequate supplies and resources for teachers in their classrooms, and offer adequate professional development. Thus, rural schools may struggle to find qualified teachers to fill the teaching positions, and when the positions are filled the teachers may be under qualified and may not stay long (Provasnik et al. 2007).

Rural schools also lack the ability to access resources because of distance and transportation limitations (Howley et al. 2001). Students may also have long bus rides and may not be able to attend before or after school activities because transportation is unavailable. Access to early childhood education and services is typically less than what other metropolitan locales have available to them. Moreover, rural schools often receive less state funding than suburban or urban districts because state-funding formulas, often based on local property tax revenues, favor larger or wealthier districts (Reeves 2003; Hass 2000). Funding is also often tied to number of students and because rural schools often have fewer students, the schools receive smaller amounts of money.

The relatively small population within specific subgroups of students (e.g., students with disabilities, English language learners, ethnicity groups) often lead to difficulties for rural schools to demonstrate adequate yearly progress because the data can be easily skewed by one or two low student test scores (Northwest Regional Educational Laboratory 2003). However, the US Department of Education (2012) allowed for increased support and flexibility for many rural schools through federal waivers and through recent developments in the Common Core State Standards Initiative. Such flexibility may be just what many experts have been saying is needed for rural school districts (Arnold 2000; Bowen and Rude 2006; Bryant 2010), but attaining high achievement standards for all students will remain difficult. The need to establish system-wide support systems and implement innovative programs to reduce the need for special education services and promote collaboration across special and regular education teachers is important. Implementing response to intervention (RTI) may be one option to better meet the needs of rural schools.

Previous Research

A review of 11 field studies implementing RTI models in rural schools was conducted to examine benefits of RTI and to consider the implications for rural special education. The studies included one or more different school settings. Researchers found that the RTI models supported slight improvements in academic performance and decreased special education referral and placement rates (Dexter et al. 2008). The outcome measures included mathematic and reading outcomes, special education referral and placement rates, and discipline and behavior referrals. However, the weak research design and procedures used within the studies call for cautious interpretation of the results. Overall, the researchers concluded that specific factors that supported scalability and sustainability of RTI programs, such as comprehensive professional development, administrative support, teacher buy-in, involving all school personnel, and adequate meeting time, were necessary regardless of setting. Professional development to support implementation of RTI in rural regions should involve the active participation of school administrators and leadership teams, a quality data management system, team planning time and networking across school sites, and publicly celebrating success (Bergstrom 2008).

Qualitative research showed positive changes occurred after the implementation of RTI within three rural schools (Shepherd and Salembier 2011). The authors identified five common factors across the school sites that appeared to promote sustainable implementation of the RTI initiative. These factors included the need to conceptualize RTI as a general education initiative, the need for effective building-based leaders, establishment of a consistent assessment practices, utilizing professional development throughout the implementation process, and creating a school-wide organizational structure that promotes active engagement of teachers and administrators in the decision-making and problem-solving process.

RTI Framework in Luverne

Luverne Elementary School is located in the rural community of Luverne, Minnesota. Luverne is in Rock County, which is in the southwest corner of the state and is primarily a farming community with some small businesses. Approximately 1177 students attend school in Luverne schools, which consists of one elementary school, one middle school, and one high school all within one central campus. A total of 540 students attend the elementary school, 88.5 % of which are white and 41 % are eligible for a free or reduced-price lunch. The elementary school has 37 licensed staff members. We also share a school social worker with the middle school and high school 3 days a week, have a school counselor on staff, and have one school psychologist for the district. There are between four and five teachers per grade level. Other licensed staff represents the special education staff along with the physical education and music staff members. The district has five Title I paraprofessionals and nine special education paraprofessionals in the elementary school.

The Model

The school district introduced the concept of RTI in 2006, and began conducting professional development workshops. In 2007, the district collected reading fluency data on students in grades two through five as a universal screener. Central administration concluded that Luverne Elementary School was ready to move forward with their RTI program in the spring of 2008, which resulted in a half-time RTI coordinator position being created. The RTI coordinator also completed the academic portions of special education evaluations.

The RTI program started small, first starting with fourth-grade groups, and then adding additional grade levels throughout the year. This was done intentionally, as a way to educate staff and show the benefits of small-group intervention for reading for all students. By the end of the school year, all grade levels were participating and staff understood the benefits of the program. During the first year, a total of 121 students were served under the RTI framework. The RTI program has continued to expand and develop into a successful program. Below is a description of our model and how we implemented it.

Grade-Level Teams

Grade-level teams served as the foundation for Luverne Elementary's RTI model. However, the teachers in the school required additional support in consuming screening data, which was provided by the building's data management team, which consisted of the RTI coordinator and the Title 1 teacher. The data management team was responsible for compiling student data into formats that grade-level teams could consume and to lead the grade-level team meetings at which the data were interpreted.

Each grade-level team met every 4–6 weeks to discuss student screening and monitoring data. Once each month, the grade-level team examined student assessment data. The meetings that immediately followed universal screening (September, January, and May) were dedicated to reviewing those data, and grade-level team interpreted progress monitoring data for the students getting interventions at the remaining meetings. One representative of the data management team (RTI coordinator and Title 1 teacher) met with each grade-level team to discuss student progress, need for change of an intervention, as well as who will be delivering interventions in grades two through five. The meetings offered the group the opportunity to discuss specific student concerns and problem-solve possible intervention changes needed to help increase student success.

Benchmarking and Data-Based Decision-Making

Students were screened three times each year to determine which students were in the bottom 20 %. Screenings were completed over a 1–2 week period. During this time, Title I teachers and the RTI coordinator pulled students individually throughout the school day to complete the screenings. Individual grade levels were screened within a 5-day period to help ensure consistent data. The data were entered into the AIMSweb database and also into a spreadsheet by those who administered the screenings. Because this was a relatively small elementary school, we also used Microsoft Excel to manage our data rather than a data warehouse. The RTI coordinator was proficient in using Excel to manage the data and helped set files up for teachers.

Once all of the data were compiled from AIMSweb and other sources, tentative groups were organized by each grade level. The grade-level team meeting then allowed teachers the opportunity to provide input based on classroom assessment data to the groupings before finalizing the lists. Changes were then made and schedules were distributed to teachers so they will know when and where interventions would occur.

One aspect of our model that was unique to rural schools was the target scores used to make decisions. Most assessment systems such as AIMSweb provide national norms, but we were not sure that those norms would be appropriate given our rural population. Therefore, target scores for determining if students were in need of intervention were taken from the St. Croix River Education District (SCRED 2012) for oral reading fluency because SCRED was also a rural district in Minnesota with a comparable population.

Small-Group Intervention

The interventions for the small groups were delivered through a combined model that used both a pullout system (i.e., RTI Time) and teacher-delivered interventions during a "Power Hour." The former was used for kindergarten and first-grade students, and the latter for older students.

RTI Time There were approximately 18 kindergarten students at any one time who needed intervention and they were served by the Title I teacher throughout the school day in small-group settings. Students were primarily placed in groups by classroom in the fall and the intervention focused on phonemic awareness and letter names/sounds. However, the groups were redistributed throughout the year as student skills and response to the intervention differentiated. The interventions were delivered to small groups of three to five students in the Title I classroom, and the sessions were 20 min in length.

First-grade students were served by a Title I teacher and the RTI coordinator during two sets of 25-min blocks. Interventions for first grade focused primarily on phonemic awareness and phonics skills using Stepping Stones to Literacy (Cooper et al. 2004) or Sound Partners (Firebaugh et al. 2005), because they focused on the appropriate skill, were research-based, easy to implement, and affordable.

Power Hour Students in second through fifth grade participated in specific 25-min time slots scheduled throughout the day that were referred to as the Power Hour (e.g., 9:30–9:55 for second grade and 10:00–20:25 for third grade). During that time, students worked on phonics, fluency, and comprehension skills depending on their specific skill deficits as determined by the results from the universal screenings. Students were identified as requiring comprehension intervention if they demonstrated adequate reading fluency, but did not score in the proficient range on the state reading accountability test. Students who scored below benchmark on the fluency measure received a fluency intervention. Students were identified as needing phonics instruction based on their accuracy (less than 93 % of word read correctly, Gickling and Thompson 1985) rate as well as their proficiency with nonsense word fluency.

One or two classroom teachers (depending on the grade), a Title I teacher, the RTI coordinator, and a paraprofessional taught the intervention groups at each grade level. Thus, there were four or five intervention groups running at each grade at any given time. The classroom teacher that led the intervention groups rotated every 6 weeks so that all teachers understood the skills that being taught, recognized the importance of the intervention, and felt a sense of ownership in the program. Groups were taught in various locations around the building. The Title I teacher and RTI coordinator utilized their individual classrooms and classroom teachers either used the back of their room, while a paraprofessional managed the rest of the students, or divided their other students between other classrooms. Group size ranged from four to six students each.

Interventions Students in need of a phonics intervention participated in one of two interventions, Sound Partners or Rewards (Archer et al. 2006). The Rewards program teaches students to break words into meaningful parts when decoding unfamiliar words and is geared toward students in fourth grade or higher. Fluency interventions were delivered with Read Naturally (2004), which is a computer-based program designed to incorporate repeated readings to increase fluency skills. Students completed placement tests and then practiced reading stories at their individual levels. A second fluency intervention was Increasing Reading Fluency through High Frequency Word Phrases (Fry and Razinski 2007), which also focuses on repeated readings, but is done in a small group with a certified teacher. Finally, Six-Minute Solution (Adams and Brown 2007) was used for fluency, which is based on a series of 1-min timings where students practice reading both word lists and stories at their individual levels. When data indicated that students were in need of a comprehension intervention to increase their critical thinking skills, then they participated in Focus on Reading Strategies (Mills 2004).

Progress Monitoring

Progress was monitored on a weekly basis for all students. Data were collected using AIMSweb probes. All students in grades one through five receiving reading interventions were administered one reading probe per week at their instructional reading level. Progress monitoring data were collected on Friday of each week by the RTI coordinator and Title I teachers.

The progress monitoring data were used to determine if students were ready to be dismissed from their RTI intervention. A formula was designed for specific grade levels to determine if they meet exit criteria. After a benchmarking period, the dismissal goal increased by one or two compared to the benchmark goal to ensure they will meet the upcoming benchmark period. Once students reached the benchmark criterion score for at least 3 weeks in a row, students were dismissed from their interventions. They were then progress monitored for 3 additional weeks to ensure that they continued to make the necessary progress to be successful with their reading skills.

Progress monitoring data were reported to teachers weekly through e-mailed spreadsheets that listed all students, their intervention, and a record of their reading results. Students who have reached their reading goal were highlighted in green so as to be easy for teachers to identify who was reaching their goals. Once students were dismissed and being monitored, the letter “M” preceded their score to denote that they were being monitored. At the bottom of each spreadsheet was a description of the exit criteria and expectations in order for all staff to better understand the spreadsheet.

Results and Lessons Learned

Since starting our research-based intervention groups for reading, Luverne Elementary School has seen progress in the reading skills of their students. Below we discuss our data and what we learned from implementing our RTI model.

Data

During the first year of interventions, we had a number of students in need of phonics interventions. As students have moved through the interventions, we have noted a decreased need for phonics interventions in the upper grade levels. Moreover, teachers have come to understand and appreciate the importance of the intervention groups and they work to make them as successful as possible.

The number of students served has increased since the inception of the program. During the 2008–2009 school year, a total of 121 students in grades kindergarten through fifth grade received interventions for reading instruction. Among those 121 students, some moved from phonics to fluency interventions and from fluency to comprehension. In the 2011–2012 school year, a total of 153 students in grades kindergarten through fifth grade received reading interventions. Students again moved to different interventions as their progress allowed.

In the spring of 2009, there were ten students in grades two through five in need of a phonics intervention. By the spring of 2012, there were only four students in need of phonics intervention. This is especially apparent when considering fourth- and fifth-grade students. When the RTI program began in the fall of 2008, there were 20 students that were inaccurate readers (i.e., reading less than 93 % of the words correctly, Gickling and Thompson 1985). By the fall of 2011, there were only two inaccurate readers in grades four and five.

The RTI program has also helped students increase their reading fluency skills. Students in fluency interventions have increased their reading fluency, and ultimately comprehension skills as they move through the program. Classroom teachers have reported increased performance in the general classrooms, in the area of reading. This can be credited, at least in part, to the progress they have made through the fluency interventions. During the first year of implementation, there were 141 (25 %) students below the 50th percentile on national norms for reading fluency and by the end of the most recent year, there were a total of 94 (16 %) students below the 50th percentile. Although the number of students at Luverne Elementary has remained mostly constant, the percentage of students eligible for the federal free or reduced price lunch program has gone up every year since implementing the RTI model in 2008.

The school's progress in implementing RTI also translated into fewer referrals for special education services because students' needs were met within the regular education setting through quality classroom instruction and research-based interventions taught during their Power Hour, or RTI times. During the first year of implementation, there were 22 students (4 % of the population) referred for special education, with 14 being diagnosed with a disability. Most recently, there were 13 students (2 % of the population) referred for initial special education testing, with only 8 requiring special education services.

Of course, the number of students who score in the proficient range on state accountability tests are the gold-standard criterion for evaluating educational reform. Luverne Elementary's students have always scored well on the Minnesota state test, but we saw a decline in scores for 3 consecutive years before beginning the RTI model. After implementing RTI, we saw a 3-consecutive-year increase in the percentage of students who passed the state test. The number of students who scored within the proficient range at Luverne Elementary fell at the 70th percentile for schools in Minnesota in 2005–2006, 58th percentile in 2006–2007, and 47th percentile in 2007–2008. That trend reversed immediately after implementing RTI as Luverne Elementary school fell at the 89th percentile compared to schools in Minnesota, which remained at about the 85th percentile in subsequent years. Moreover, the Minnesota Rural Education Association presented Luverne Elementary School with the 2011 Profiles of Excellence Award of Distinction for the Elementary RTI Program.

Considerations for Rural Schools

Although many of the core elements of our model were consistent with most approaches, there were some considerations for our rural school. First, when implementing RTI in a rural setting, it is important to be flexible. School personnel should examine all possible staff members available to teach interventions, especially given that rural schools may often have less support personnel than more urban or suburban schools. Given the relative lack of resources, school personnel in rural schools should determine resources available and start small.

Second, many rural schools face unique challenges with staff recruitment and retention due to limited financial resources, lower pay as compared to urban and suburban schools, and relatively inadequate supplies and resources for teachers in their classrooms (Barley and Beesley 2007; Monk 2007). This often creates a situation in which rural areas cannot find qualified teachers to fill the teaching positions, and when the positions are filled the teachers may be under qualified and may leave the district within the first 5 years (Provasnik et al. 2007). Thus, it is essential that teachers buy into the program for it to be successful and that sufficient training occurs for new teachers. We found that having a set list of possible interventions was helpful in training and in generating interest among teachers, but it was critically important to have a culture of data-based decision-making as led by the school principal. We also used the Institute for Education Science Practice Guide regarding RTI for elementary schools (Gersten et al. 2009) as a useful training resource for grade-level teams. The major recommendations from the document and how we implemented it are included in Table 1.

Third, rural schools often receive less state funding than suburban or urban districts because state-funding formulas, often based on local property tax revenues, favor larger or wealthier districts (Reeves 2003; Hass 2000). Funding is also often tied to number of students and because rural schools often have fewer students, the schools receive smaller amounts of money. Therefore, school personnel need to adopt a model that focuses on efficient use of resources and does not require additional personnel or resources to implement.

Finally, rural schools struggle to provide appropriate educational and mental health services to all students but especially those identified with disabilities (Williams 2010). For example, delivering psychological services in rural communities is often challenging because of the smaller number of psychologists, counselors, social workers, and psychiatrists available within these communities in comparison to metropolitan areas (Richgels and Sande 2009). Additionally, rural schools struggle to recruit and retain special education teachers and a shortage of qualified personnel to serve special education students remains a major problem in rural areas (Gehrke and McCoy 2007; Hodge and Krumm 2009). Therefore, rural schools need to focus on malleable factors for students who struggle academically. School personnel who see student difficulties as a disability in need of specialized instruction may unwittingly confound a potentially difficult situation. Alternatively, school personnel should analyze student difficulties in great detail to find the environmental variable within the school's control

Table 1 Summary table of implications for practice

Research-based recommendations from Gersten et al. (2009)	Practices for implementation
`	Instruction is focused on student need and students moved to different groups as their needs change
	Kindergarten focuses on phonemic awareness and phonics
	First grade focuses on phonemic awareness and phonics and moves into fluency for the second half of the year, as appropriate
	Second through fifth grade focuses on phonics, fluency, and comprehension based on student need
	Tier 2 intervention is implemented in grade kindergarten through fifth grade
Provide instruction in small groups	Interventions are implemented in small groups of 2–6 students
Typically, these groups meet 3–5 times per week for 20–40 min	Interventions are implemented four times each week
	Interventions are 20 min for kindergarten
	Interventions are 25 min in grade one through five

that is most directly related to the problem (Burns and Gibbons 2012). Taking a more environmental interpretation of student difficulties will likely result in more efficient use of resources, in-depth analysis to better understand student difficulties, and a belief among all teachers that all students can learn, which is especially important in a rural school.

Directions for Future Research

Although we were able to implement an RTI model that led to positive outcomes for our students, there were several questions to which the answer was not apparent and requires additional research. First, we identified research-based interventions that led to positive results, but additional research is needed regarding the effectiveness of specific intervention programs and how they can most efficiently be implemented in schools with limited resources. While it is important not to limit interventions allowed for RTI purposes, it would be helpful if districts knew which interventions would yield the highest results and which can be implemented on a small budget.

Another possible area of further research would be the long-term effect of RTI on student performance for students who attend rural schools. Are students able to maintain their increased reading skills and does that translate into higher academic achievement in future grade levels and beyond high school? Moreover, does implementing an RTI model in a rural school affect teacher recruitment, retention, and satisfaction?

A final potential area for research could be the effect that implementing an RTI model may have on special education within a rural school. Many rural schools do not have the same amount of special education services as do many urban or suburban schools and, therefore, must be more creative in meeting the needs of unique learners. We saw a decrease in the number of children referred to and receiving special education services. Thus, it may be possible for special education personnel to be freed up to provide more prevention-focused services or to meet the needs of children with more intense disabilities.

Conclusion

In conclusion, the adoption of RTI in the rural setting can potentially have positive results on student performance. When any school is able to target specific needs of students, progress will be made. It is important to get started with the process, no matter the demographics of your district, but it seems especially relevant to meet the unique challenges of rural schools. When small-group instruction is organized and delivered based on student need as determined by data, all students will benefit. However, it is essential to help all staff members take ownership in the program, to provide sufficient support, and to systematize the process so that the level of implementation does not change when staff turns over. Most importantly, schools need to take the first step: start the process.

References

A practice guide. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Adams, G., & Brown, S. (2007). *The six-minute solution: A reading fluency program*. Longmont: Sopris West.

Archer, A. L., Gleason, M. M., & Vachon, V. (2006). *REWARDS: Multisyllabic word reading strategies*. Boston: Sopris West.

Arnold, M. L. (2000). *Rural schools: Diverse needs call for flexible policies*. Aurora: Mid-continent Research for Education and Learning.

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., & Zhang, J. (2013). *The Condition of Education 2013* (NCES 2013-037). Washington, DC: U.S. Department of education, national center for education statistics. http://nces.ed.gov/pubsearch.

Barley, Z. A., & Beesley, A. D. (2007). Rural school success: What can we learn? *Journal of Research in Rural Education, 22,* 1–16.

Bergstrom, M. K. (2008). Professional development in response to intervention: Implementation of a model in a rural region. *Rural Special Education Quarterly, 27*(4), 26–36.

Bickel, R., & Howley, C.B. (2000). *When it comes to schooling…small works: School size, poverty and stu-*

dent achievement. Randolf: Rural School and Community Trust. http://www.eric.ed.gov/.

Bowen, S. K., & Rude, H. A. (2006). Assessment and students with disabilities: Issues and challenges with educational reform. *Rural Special Education Quarterly, 25*(3), 24–30.

Bryant, J. A. (2010). Dismantling rural stereotypes. *Educational Leadership, 68,* 54–58.

Burns, M. K., & Gibbons, K. (2012). *Response to intervention implementation in elementary and secondary schools: Procedures to assure scientific-based practices* (2nd ed.). New York: Routledge.

Cooper, P., Gonzalez, J., & Nelson, J. R. (2004). *Stepping stones to literacy*. Longmont: Sopris West.

Dexter, D. D., Hughes, C. A., & Farmer, T. W. (2008). Responsiveness to intervention: A review of field studies and implications for rural special education. *Rural Special Education Quarterly, 27*(4), 3–9.

Firebaugh, M., Jenkins, J., O'Connor, R., Peyton, J., Vadasy, P., & Wayne, S. (2005). *Sound partners*. Longmont: Sopris West.

Fry, E., & Rasinski, T. (2007). *Increasing reading fluency with high frequency word phrases*. Huntington Beach: Shell Education.

Gehrke, R. S., & McCoy, K. (2007). Sustaining and retaining beginning special educators: It takes a village. *Teaching and Teacher Education, 23,* 490–500.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, D. W. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades*. Washington, DC: Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance.

Gickling, E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 205–218.

Hass, T. (2000). Balance due: Increasing financial resources for small rural schools. http://www.ericdigests.org/2001-3/due.htm.

Hodge, C. L., & Krumm, B. L. (2009). NCLB: A study of its effects on rural schools—School administrators rate service options for students with disabilities. *Rural Special Education Quarterly*, *28*(1), 20–27.

Howley, A., Howley, C., & Shamblen, S. (2001). Riding the school bus: A comparison of the rural and suburban experiences in five states. *Journal of Research in Rural Education, 17*(1), 41–63.

Johnson, J. (2004). *Small works in Nebraska: How poverty and the size of the school systems affect school performance in Nebraska*. Randolf: Rural Schools and Community Trust.

Keaton, P. (2012). *Numbers and types of public elementary and secondary schools from the Common Core of data: School year 2010–11* (NCES 2012-325rev). U.S. Department of Education. Washington, DC: National Center for Education.

Mills, S. (2004). *Focus on reading strategies*. Logan: Perfection Learning.

Monk, D. (2007). Recruiting and retaining high-quality teachers in rural areas. *The Future of Children, 17,* 155–174.

National Center for Education Statistics. (2010). Status and trends in the education of racial and ethnic minorities. http://nces.ed.gov/pubs2010/2010015/tables/table_7_2.asp.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.

Northwest Regional Educational Laboratory. (2003). Challenges and opportunities of NCLB for small, rural, and isolated schools. http://oregonpirc.org/?q=webfm_send/23.

Pearson, Inc. (2013). *AIMSweb*. Bloomington: author.

Provasnik, S., KewalRamani, A., Coleman, M. M., Gilbertson, L, Herring, W., & Xie, Q. (2007). *Status of education in rural America*. Washington, DC: National Center for Education Statistics.

Reeves, C. (2003). *Implementing no child left behind act: Implications for rural schools and districts*. Medford: North Central Regional Educational Laboratory.

Richgels, J., & Sande, C. (2009). Collaboration is key for rural challenges. *Journal of Policy and Practice, 67,* 17–19.

Shepherd, K. S., & Salembier, G. (2011). Improving schools through a response to intervention approach: A cross-case analysis of three rural schools. *Rural Special Education Quarterly, 30*(3), 3–15.

St. Croix River Education District. (2012). *Target and norm charts*. Rush City: St. Croix River Education District. http://www.scred.k12.mn.us/UserFiles/Servers/Server_3022443/File/Research%20&%20Outcomes/2012-2013%20SCRED%20Target%20Packet.pdf.

Strange, M., Johnson, J., Showalter, D., & Klein, R. (2012). Why rural matters 2011–2012: The condition of rural education in the 50 states. A report by the Rural School and Community Trust. www.ruraledu.org.

US Department of Education. (2012). *ESEA flexibility*. Washington, DC: author. http://www.ed.gov/esea/flexibility.

Williams, D. (2010). *The rural solution: How community schools can reinvigorate rural education*. Washington, DC: Center for American Progress.

School-wide Positive Behavior Support and Response to Intervention: System Similarities, Distinctions, and Research to Date at the Universal Level of Support

Timothy J. Lewis, Barbara S. Mitchell,
D. Tichelle Bruntmeyer and George Sugai

The logic of early intervention and prevention that pervades all of the chapters within this text was first used within the public health arena (Walker and Bullis 1991; Walker et al. 1996). The basic concept was to first prevent a specific health challenge, for example, skin cancer, through universal supports designed to reduce the overall incidence of cases (e.g., use of sun screen, hats). Secondary prevention/intervention is aimed at catching the first signs of the problem and implementing treatment right away (e.g., routine screening of skin moles and removing precancerous cells). Tertiary prevention focuses on stopping the disease and preventing other related health issues (e.g., chemotherapy to stop the spread of melanoma to other types of related cancers). Walker et al. (1996) were the first to apply the public health logic model to address high-risk students on a pathway to developing significant behavioral challenges.

Extending the work of Walker and colleagues beyond a specific focus on conduct disorders, School-wide Positive Behavior Supports (SW-PBS) was not only designed to address the continuum of social and emotional behavioral problems in school and extend the public health logic of both preventing challenging behavioral problems in school but also creating environments that increase the likelihood of those students identified with disabilities who experience success in mainstream instructional settings (Colvin et al. 1993; Lewis and Sugai 1999; Horner and Sugai 2005). While there are several similarities between response to intervention (RTI) and SW-PBS, such as data-based decision-making, building a continuum of supports, and working with teams of educators, there are also several distinctions that should be noted. First, unlike many academic behaviors that are occasioned by specific prompts, such as reading assigned texts or materials, and expected only when those prompts are present (i.e., child is directed to read text), educators "expect" students to "behave" across all school settings and under all possible conditions throughout the day (e.g., whole-class instruction, one-on-one, structured time, unstructured time, during free play). Second, for the majority of academic challenges, such as a student struggling to learn to read, the academic challenge typically has minimal impact on the overall learning environment. In other words, unlike the child who displays a high rate of acting out and disruptive behavior such as yelling profanities and throwing objects, the child who is struggling to read successfully a paragraph during silent reading will have little impact on other students' learning at that moment. Finally, school systems have long been set up whereby educators understand

T. J. Lewis (✉) · B. S. Mitchell · D. T. Bruntmeyer
University of Missouri, Columbia, USA
e-mail: lewistj@missouri.edu

G. Sugai
University of Connecticut, Storrs, USA

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention*, DOI 10.1007/978-1-4899-7568-3_40

the logic of increasing environmental supports to increase the likelihood of academic success. If an elementary teacher is asked what he or she would do if a student was struggling to learn to read, the teacher will most likely provide a long list of increasingly intensive and individualized instructional supports (e.g., one-on-one instruction, peer tutors, supplemental practice) along with a rich schedule of positive performance feedback recognizing any progress. Ask that same teacher what he or she would do if a student was disruptive in class, verbally aggressive, or simply noncompliant and the response will not follow the logic of increasing supports, rather, the typical focus will be on removing that child from the classroom (Bradley et al. 2004).

In this chapter, and the related chapter on small-group and individual behavioral supports, the emphasis on building strong systems whereby adult needs are given equal attention is evident and critical for success. In other words, an additional hallmark of SW-PBS is the great amount of attention given to educator supports recognizing that (a) most general educators and administrators have not received extensive training in how to provide social and emotional behavioral supports, (b) current school "discipline" systems are set up to remove students who present problems, not to "push in" supports to keep the child in the classroom, and (c) given the expectation that students display appropriate behavior from the moment they enter to the moment they exit the school building, success will require schoolwide systems that include all adults within the building. Therefore, the purpose of this chapter is to provide an overview of essential features of SW-PBS and review efficacy research conducted to date. This chapter provides an overview of the challenges students at risk for emotional behavioral problems present within educational settings, challenges all students face in schools with ineffective behavioral supports, and the essential environmental supports necessary to increase successful outcomes. Implications for research and practice are also discussed.

Behavioral Challenges in Schools

Recent episodes of extreme violence have prompted the reemergence of school safety as a top priority, and the current data on school crime provide evidence of this pressing need. For example, 85 % of public schools recorded one or more crime incidents that took place during the school year 2009–2010 and 74 % recorded one or more crimes that were violent (Robers et al. 2013). During the school year 2010–2011 there were 31 school-associated violent deaths, and approximately 1.2 million nonfatal victimizations, ranging from simple to serious assault, in school settings (Robers et al. 2013). Further during 2011, among youth aged 12–18 years, a greater number of students experienced crime at school (i.e., theft and violence) than away from school, representing an increase from the previous year (Robers et al. 2013). Data from students in grades 9 through 12 are particularly alarming, showing 7 % of adolescents reported being threatened or injured with a weapon (e.g., gun, knife, or club) while on school property (Robers et al. 2013). Beyond the events of crime and violence, bullying behavior, acts of disrespect toward teachers, student racial/ethnic tensions, student sexual harassment of other students, gang activities, and widespread disorder in classrooms are also documented as the most frequently occurring problem behaviors in school settings (Robers et al. 2013). In response to these challenges, evidence also suggests an increase in the security measures schools enact. In 2011, 95 % of students reported requirements for school visitors to sign in upon arrival, 77 % reported the use of security cameras, and 70 % reported security guards and/or police officers working in their schools (Robers et al. 2013).

While the majority of students will not experience exceedingly violent or aggressive events, student problem behavior consistently has been reported as one of the top concerns among educators (Rose and Gallup 2007; U.S. Department of Education 1998). Administrators and teachers report that responding to issues of school discipline was one of the greatest demands of their time, cite problem behavior as interfering with instruction and learning, and report disruptive behavior as the most common reason for removal of students from classroom or school settings (Miller-Richter et al. 2012). For example, during the school year 2009–2010, 39% of public schools reported taking a serious disciplinary action against a student (Robers et al. 2013). Of the 433,800 serious disciplinary actions taken that year, 74% were multiday suspensions (i.e., 5 days or more), 20% included transfer of students to a specialized school, and 6% used removal from school, without access to services, for the remainder of the school year (Robers et al. 2013). Even more alarming, the fastest growing rate of suspensions and expulsions due to challenging behavior are occurring at the preschool level (Gilliam and Shabar 2006).

Undeniably, ensuring school safety and establishing a positive and productive learning environment is of utmost importance. At the same time, schools continue to be confronted with an ever-increasing complexity of student needs, above and beyond academic instruction. Prevalence estimates indicate that approximately 20% of the school age population currently experience a mental, emotional, or behavioral disorder with a majority of conditions emerging during the early years of learning (e.g., median age of onset age 6 for anxiety, 11 for behavior, 13 for mood, and 15 for substance abuse disorders; Merikangas et al. 2010). Yet, less than 1% of all students are identified as having an emotional/behavioral disorder (E/BD) and determined eligible to receive services within school settings (U.S. Department of Education, Office of Special Education Programs, 2011).

It is well documented that children and adolescents with mental, emotional, and/or behavioral challenges are at great risk for a host of increasingly negative outcomes that include: (a) inadequate engagement with school and learning reflected by poor attendance and difficulties establishing or maintaining appropriate relationships with teachers and peers (Lane et al. 2006; Merrell and Walker 2004; Wagner et al. 2005); (b) lower academic achievement than any other category of students with disabilities (U.S. Department of Education 2008); (c) limited graduation rates with more than half dropping out before completion (U.S. Department of Education 2004); and (d) increased risk for incarceration, substance abuse, unemployment, and suicide (Wagner et al. 2005). Related risk factors such as poverty, family disruption, child abuse or neglect, ineffective peer relationships, and community violence that contribute to poor psychological and school adjustment also have been identified clearly (National Research Council and Institute of Medicine (NRC & IOM) 2009).

Fortunately for educators and school support personnel, interventions for promoting social and emotional well-being, preventing the transition from "risk" to disorder, and lessening the intensity of impact of a disability are available (Lewis et al. 2010; Peacock Hill Working Group 1991; NRC & IOM 2009). However, current special education procedures for identification and intervention continue to be reactive rather than preventive (Gresham 2007; Maag and Katsiyannis 2008). For example, IDEA eligibility criteria require that students demonstrate patterns of problematic behavior to a "marked degree" and "over a long period of time" before an evaluation can be completed, diminishing the window of opportunity for early and successful intervention (Code of Federal Regulations, Title 34, Section 300.7(c)(4)(i)).

In summary, educators are faced with several behavioral challenges in schools. First, schools are confronting increasingly higher rates of aggressive and violent behavior and yet continue to rely on reactive strategies that largely involve exclusion, which has been documented as ineffective among high-risk students (Walker et al. 2004). Second, educators often are the first responders to significant mental and emotional student issues, but are poorly equipped to intervene and install supports across school settings. Finally, when educators exhaust what limited strategies they may have at their disposal and suspect a possible disability, the special education evaluation process itself necessitates that educators document student failure over time and to a "marked degree" as a necessary component for eligibility under the category of EBD. The culmination of a slow-to-respond system of specialized supports, exclusion as the most frequently used intervention option, and environments unprepared to promote good mental and emotional health has resulted in the continued high-failure rate of students with, and at risk for, EBD (Bradley et al. 2004; Wagner et al. 2005).

Features of School-wide Positive Behavior Support

Similar to RTI, SW-PBS is best characterized as a problem-solving framework; it is not a curriculum, program, or intervention package (see the *Implementer's Blueprint* for a comprehensive overview of essential SW-PBS components; pbis.org). First, schools assemble a representative team that is tasked with development, implementation oversight, and maintenance and generalization of the SW-PBS process. The team must include an administrator and a cross section of faculty and staff within the building (e.g., grade level, specialists, staff). The first step in the SW-PBS process is to review existing data sources and complete a systems assessment to determine current behavioral challenges and the school's capacity to address the noted challenges (see Fig. 1). Data sources include behavioral and discipline infractions, in- and out-of-school suspensions, attendance, achievement, time out of instruction, and any other relevant archival data. Self-assessment tools such as the *Self Assessment Survey* (SAS; pbis.org) allow teams to evaluate what behavioral systems from school-wide to individual student support are currently in place and the degree to which the current systems are responsive to the unique challenges of the school.

Based on their data, SW-PBS teams identify evidence-based practices to put in place at the school-wide, nonclassroom (e.g., hallway, cafeteria), classroom, and individual student level. In concert with intervention selection, ongoing data sources are also identified to monitor intervention outcomes (see Fig. 1). Unlike traditional "packaged" programs, SW-PBS emphasizes a problem-solving framework to assist school teams in the selection of evidence-based behav-

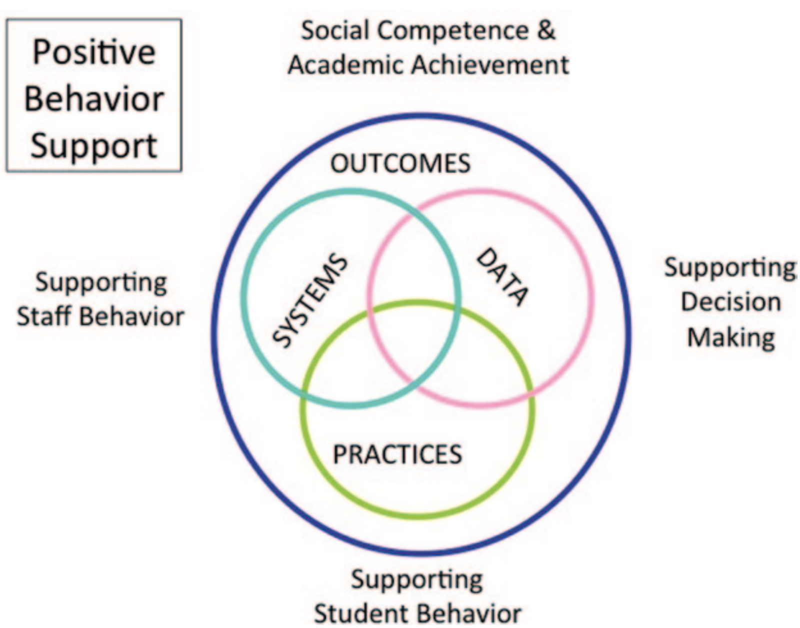


Fig. 1 School-wide positive behavior support (SW-PBS) problem-solving logic model. (Available from the Office of special education programs center on positive behavior interventions and supports; http://pbis.org)

ioral interventions and supports, matched to the presenting problems. In addition to clear empirical evidence demonstrating student outcome improvements for certain intervention practices, advocated practices within a SW-PBS framework build on an instructional model to capitalize on the common strength found across school staff; that is, teachers are recognized to have the ability to teach, provide practice opportunities, and provide specific feedback on skill use, including adaptive behavior skills (Sugai et al. 2000). Instructional targets within SW-PBS include (a) pro-social alternatives to noted problem behaviors, (b) specific behavioral challenges such as bullying, (c) cognitive strategies to address internalizing concerns such as anxiety, social-emotional development, and (d) overall physical and mental health and well-being. Key to successful implementation of SW-PBS practices is to ensure all staff within the school are fluent in implementation of selected strategies to promote high-implementation fidelity using operationalized training and ongoing technical assistance that includes performance feedback.

The final key element of SW-PBS is a clear and concerted focus on the necessary system supports to ensure fidelity of intervention implementation and accurate data collection for progress monitoring (see Fig. 1). The focus of SW-PBS is on creating environments to increase the likelihood students engage in appropriate social behavior across all school settings. Environments that increase the likelihood of appropriate social behavior are guided by a core curriculum that is implemented with consistency and fidelity that reflects the unique behavioral challenges and social context of the school (Lewis 2010). Therefore an essential task for the school team is to develop and implement training and technical assistance to assist all school personnel to successfully implement the social-behavioral curriculum and environmental supports. While the focus of SW-PBS is on student outcomes, the majority of the school team's time will be dedicated to supporting their colleague's implementation of practices and accompanying environmental supports (e.g., prompting skill use, providing specific feedback on correct and incorrect responding). Building on best practices noted within the professional development literature (e.g., Fixsen et al. 2005; Guskey 2000), SW-PBS through a network of school, district, regional, and state supports emphasizes the following key components in all professional development activities (see *Professional Development Blueprint*, pbis.org for more details):

- Readiness and prerequisite skills are addressed prior to training or technical assistance.
- Training and technical assistance is tailored to meet team needs.
- Long-term skill-based training is implemented with clear and measurable outcomes.
- Practice opportunities with coaching and performance feedback are provided.
- Professional networks are established to share strategies.

Focusing the necessary supports to promote systemic implementation across educators is critical to success and has led to the development of several fidelity measures including the *School-wide Evaluation Tool* (SET; Horner et al. 2004) and the *Benchmarks of Quality* (BoQ; Kincaid et al. 2010) measuring universal implementation, the *Benchmarks for Advanced Tiers* (BAT; Anderson et al. 2011) and the *Individual Student Systems Evaluation Tool* (ISSET; Anderson et al. 2011) which is appropriate for tiers 2 and 3.

SW-PBS, like RTI, also focuses on building an interconnected continuum of student supports through a multi-tiered continuum of increasing behavioral and academic supports (see Fig. 2). Initially, school teams use the problem-solving logic of SW-PBS (i.e., *data*-based decisions leading to the identification of *practice* and ensuring personnel can implement through *system* supports) to build universal supports designed to address the needs of all students across all school settings. Unlike academic multitiered systems of support whereby teams often adopt commercially available curricula (e.g., reading and math series) for tier 1, school teams must develop their social behavioral tier 1 program based on current problems and issues occurring in the school. Initially, school teams identify common and persistent behavior problems within their school (e.g., "noncompliance," "disrespect," "verbal aggres-

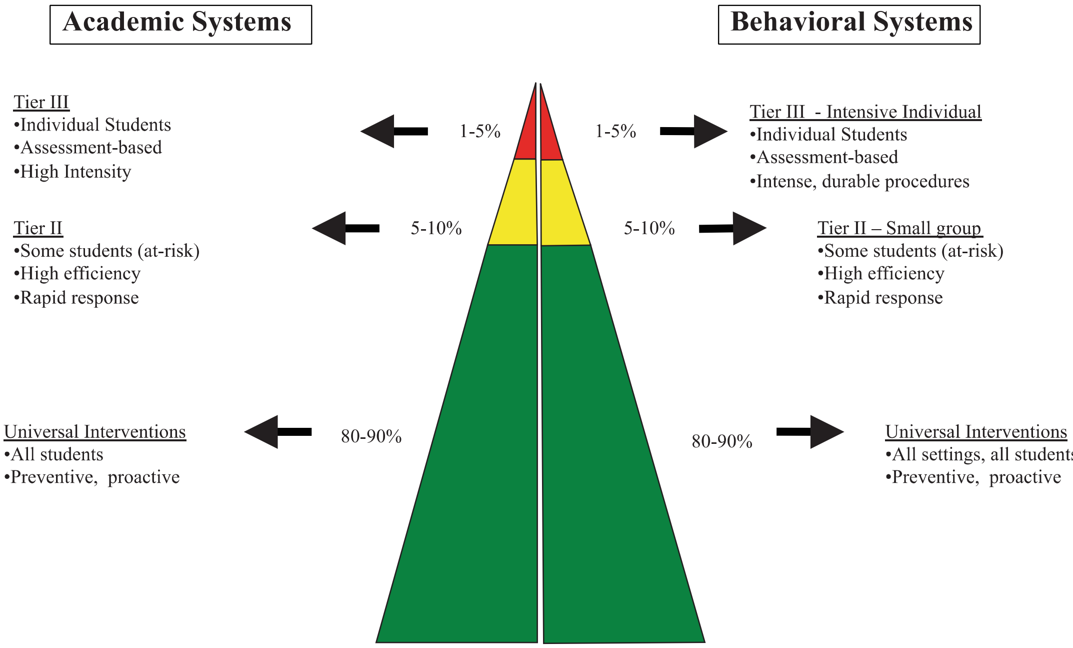


Fig. 2 The continuum of school wide positive behavior supports and response to intervention (RTI) academic supports. (Available from the Office of special education programs center on positive behavior interventions and supports; http://pbis.org)

sion," "bullying") using multiple data sources. For each problem identified, school teams define "replacement" pro-social skills. For every problem behavior, teams identify what students should "do instead." SW-PBS avoids common "codes of conduct" or "discipline policies" that outline what students should not do and the consequences if they are caught engaging in any of the behaviors. In relation to identifying specific replacement behaviors, school teams identify themes across replacement behaviors that are then translated into three to five broad rules (e.g., be respectful, be responsible, be a learner). Teams are tasked to identify replacement behaviors across all school settings, and then further identify specifics within key school settings for each broad rule (e.g., hallways, bathrooms, classrooms). The result is a matrix of desired school-based social skills that serves as that school's social behavior "scope and sequence" (see Fig. 3 for a sample SW-PBS matrix).

Once behavioral expectations are identified, the school is tasked with developing developmentally appropriate lesson plans to teach expectations, practice opportunities, and strategies to deliver positive-specific feedback when students display appropriate behavior. Social expectations are taught across the school year, follow effective instructional practices similar to academics, and are tailored to reflect student phases of learning (i.e., acquisition to fluency to maintenance and generalization). In addition to direct instruction of the locally developed core curriculum for social behaviors, school teams are also taught to view social behavior "learning errors" in the same vein as academics. Viewing less than perfect performance of desired social behaviors as an opportunity for learning requires a substantial shift among many adults in schools. SW-PBS encourages adults to move away from attempting to punish problem behavior and instead implement additional instructional and environmental support strategies to increase the likelihood of student mastery. When a student does not display the desired social behavior or exhibits disruptive behavior, this is viewed as an opportunity for adults in the student's environment to troubleshoot how the environment may have failed the

I am….	All Settings	Classroom	Hallways	Cafeteria	Bathrooms	Playground	Assemblies
Safe	• Keep bodies calm in line • Report any problems • Ask permission to leave any setting	• Maintain personal space	• Walk • Stay to the right on stairs • Banisters are for hands	• Walk • Push in chairs • Place trash in trash can	• Wash hands with soap and water • Keep water in the sink • One person per stall	• Use equipment for intended purpose • Wood chips are for the ground • Participate in school approved games only • Stay in approved areas • Keep body to self	• Walk • Enter and exit gym in an orderly manner
Respectful	• Treat others the way you want to be treated • Be an active listener • Follow adult direction(s) • Use polite language • Help keep the school orderly	• Be honest • Take care of yourself	• Walk quietly so others can continue learning	• Eat only your food • Use a peaceful voice	• Allow for privacy of others • Clean up after self	• Line up at first signal • Invite others who want to join in • Enter and exit building peacefully • Share materials • Use polite language	• Be an active listener • Applaud appropriately to show appreciation
A Learner	• Be an active participant • Give full effort • Be a team player • Do your job	• Be a risk taker • Be prepared • Make good choices	• Return to class promptly	• Use proper manners • Leave when adult excuses	• Follow bathroom procedures • Return to class promptly	• Be a problem solver • Learn new games and activities	• Raise your hand to share • Keep comments and questions on topic

Fig. 3 Sample matrix of social behavior expectations

child, so adjustments can be made to prevent the problem behavior in the future similar to a student making a math or reading "error."

Once the problem-solving logic of SW-PBS is implemented at the universal level and schools have effectively taught appropriate social behavior, have provided multiple opportunities for students to practice these new behaviors, and environmental supports have been adjusted to increase the likelihood of student success (e.g., increased proactive supervision, delivery of high rates of positive-specific praise or feedback acknowledging student mastery), school teams begin to implement additional tiers of support for students who are unsuccessful at tier 1. Tier 2 or small-group supports are designed for students who are not successful with universal supports alone. School teams develop data-based decision rules and other strategies including screening and teacher referral to catch "non-responders" early and prevent disruptive behaviors from becoming chronic and intense (see Chap. 31 for an overview of Tier 2/3 SW-PBS). Tier 2 strategies within an SW-PBS framework include additional small-group social skill instruction, self-management, and academic supports. For those students who do not respond to tier 2 supports, or in instances where problem behaviors are intensive and chronic, tier 3 or individualized supports are implemented. At the tier 3 level, a functional behavioral assessment is conducted to design an instruction-based behavioral intervention plan. Community, mental health, and specialized instructional supports and related services are also included at the tier 3 level when indicated.

At each tier, the central features of SW-PBS are repeated: (a) identify the problem/concern through data, (b) identify the desired pro-social replacement skill, (c) explicitly teach the skill, and (d) alter the environment to build in necessary supports to increase the likelihood of student success. As students require increasing intensive

Table 1 Phases of implementation of SW-PBS within each tier of support[a]

Focus	Stage	Description
Should we do it?	Exploration/adoption	School/district commits to adopting SW-PBS and building systems to support implementation
Work to do it right	Installation	School/district sets up infrastructure to implement SW-PBS (e.g., forms leadership team, conducts assessment, creates action plan)
	Initial implementation	Team starts in a specific school setting or focuses on key behavioral targets, implements evidence-based practices and develops supporting systems
Work to do it better	Elaboration	Team expands practices and systems to develop a comprehensive tier of support based on a continuous review of student outcome and fidelity of practice data
	Continuous improvement/ regeneration	Team reviews multiple data sources, eliminates inefficiencies and expands effective strategies within the tiered level of support

[a] Based on work by Fixsen et al. 2005 and Goodman 2013

supports to be successful, often staff will require a higher level of training and technical assistance to implement more intensive supports and interventions. Similar to academic systems of RTI, one key to system success is a school- and district-wide emphasis and focus on SW-PBS, aligning training and technical assistance needs to local school data. The quality of core or universal instruction influences the percentage of students who will experience success without additional intervention, and systematically influences the effects of increasingly intensive layers of intervention (Algozzine et al. 2011; Bradshaw et al. 2010; Putnam et al. 2002). An equally important component in building a continuum of student supports for academic and social behavior is carefully monitoring school readiness features with respect to each phase of implementation (Fixsen et al. 2005). In the initial implementation of universal supports, schools move through a predictable set of steps or phases from exploration to adaptation (see Table 1). Likewise, once schools reach mastery and implement universal strategies with high fidelity (e.g., reach 80% or better on the *School-wide Evaluation Tool*), they once again start at the initial phase at the start of tier 2 implementation and again at tier 3 implementation. Similar to the development of universal supports, descriptive data indicate the development and full implementation of subsequent tiers takes 1–2 years.

SW-PBS as Early Intervention/ Prevention

To date, the majority of development, research and evaluation efforts of SW-PBS have been conducted at the universal level with an eye toward implementation of previously validated practices designed to prevent problem behavior and increase the likelihood of students with disabilities and those at high risk experience success (Lewis et al. 2010). In addition, as described above, it is not sufficient to provide limited or pull-out exposure to evidence-based practices, the entire school environment should be redesigned to create a seamless continuum of supports. Such an approach necessitates that educators change their views and practices regarding the prevention and management of disruptive behavior in schools. Over 20 years ago a group of distinguished scholars in the field of Emotional/Behavioral Disorders put forth a compendium of best practices, along with the necessary environmental supports, which unfortunately were never fully implemented in the years since due in large part to absence of systemic changes in how services are delivered in schools (Peacock Hill Working Group 1991). SW-PBS, along with RTI on the academic side of the continuum, provides educators with the framework to implement best practices and create instructional environments to fully implement evidence-based practices through comprehensive system changes in how education is delivered across general and special education.

Over the past 15 years, a clear body of research has emerged documenting the prevention/early intervention effect of universal supports. Quasi-experimental and descriptive studies have shown schools can: (a) effectively and efficiently reduce the overall rates of problem behavior from preschool to high school (Barrett et al. 2008; Bohanon et al. 2006; Chapman and Hofweber 2000; Curtis et al. 2010; Duda et al. 2004; Farkas et al. 2012; Lohrmann-O'Rourke et al. 2000; Nelson et al. 1998; Putnam et al. 2002; Simonson et al. 2010), (b) improve academic outcomes through improvements in behavioral supports (Algozzine et al. 2011; Luiselli et al. 2005; McIntosh et al. 2006; 2008a, b, 2012), and (c) improve classroom and non-classroom outcomes by targeting specific SW-PBS strategies in those settings (De Pry and Sugai 2002; Hirsch et al. 2004; Lewis et al. 2000, 2002; Putnam et al. 2003; Stichter et al. 2006). Several recently conducted randomized controlled trial studies have confirmed previous outcomes including positive sustained changes in school discipline practices that result in decreases in problem behavior and increases in appropriate behavior (Bradshaw et al. 2008b, 2010, Horner et al. 2009), overall school climate improvements (Bradshaw et al. 2008a, 2009), and the reduction of specific behavioral challenges (Bradshaw et al. in press; Waasdorp et al. 2012) with moderate effect sizes across each targeted outcome.

SW-PBS as Behavioral RTI Within the Special Education Eligibility Process

The development and evaluation of SW-PBS predates the recent reauthorization of IDEA that allowed the use of academic RTI as one facet of eligibility determination within the category of Specific Learning Disabilities (LD; IDEA 2004). While the major focus of SW-PBS has been on prevention, early intervention, and altering instructional environments to maximize behavioral intervention effectiveness, recent work has called for a move from the largely medical model of evaluation to determine student eligibility within the category of "Seriously Emotionally Disturbed" (SED; Lewis et al. 2010; Maag and Katsiyannis 2008; Merrell and Walker 2004; Mathur 2007). Over two decades ago, a group of education, mental health, and other related professionals proposed the first significant changes to the SED definition and evaluation process. These scholars advocated for a broader evaluation process that examined behavioral functioning across multiple settings in addition to reviewing data on how students responded to behavioral supports suitable for implementation in the general education environment (Forness and Knitzer 1992).

Unfortunately, the proposed EBD definition and evaluation process that was better suited to an instructional or educational framework did not make it into the reauthorization of IDEA in 2004. However, recent calls to consider response or "non-response to interventions" implemented with fidelity, in addition to other evaluation data sources (e.g., rating scales, academic testing) and evaluation processes (e.g., ruling out possible causes), continues to be advocated in the professional literature especially in light of the under-identification of children and youth within the SED category (Cheney et al. 2008; Fairbanks et al. 2007; Gresham 2007; Hawken et al. 2008; Lewis et al. 2010; Maag and Katsiyannis 2008). As stated at the outset of this chapter, while the parallel process of examining nonresponse data to intervention within the context of comprehensive multi-tiered systems of support as one source of evidence of a possible SED, similar to the present work in SLD, has the potential to provide individualized more intensive supports and avoid the current "wait to fail" model. However, the complex interactive nature of social behavior across all school settings does not lend itself to a quick benchmark measure that can be monitored across a targeted intervention within one or two academic settings or conditions. Rather, the knowledge-base at this point suggests that nonresponse data should be viewed as one component of the overall evaluation process necessitating that additional data sources, such as parent interviews or teacher-rating scales, should continue to be an essential component of the evaluation process (Kauffman et al. 2009). In addition, student nonresponse data to interventions as a facet of the special education eligibility process will be valuable only to the degree the evaluation team has confidence the prior interventions were: (a) evidence-based with clear demonstration of effect on other children or youth with emotional or behavior problems, (b) matched to student need based on a clear data-based process, (c) were implemented with integrity over a sufficient period of time, and (d) the progress monitoring data were clearly operationally defined and collected consistently across all staff in the school. At present, these criteria are not routinely met in the majority of schools in the USA. At present, a hybrid of social-behavioral RTI combined with traditional evaluation data (e.g., rating scales, archival review, direct observations) is warranted in the determination of SED (Gresham 2005, 2007; Kauffman et al. 2009; Lewis et al. 2010).

Implications for Research

While the emerging evidence base of descriptive, quasi-experimental, and experimental research has provided a solid knowledge base on the impact of SW-PBS universal supports, several lines of research are still needed. First, while overall impact on behavior across all students has been well documented, it is unknown what impact universal programs and the complete continuum of multi-tiered supports have on high-risk students with respect to maintenance and generalization of student behavior change. While the component practices within SW-PBS have a long history of demonstrating student outcomes, the added value of implementation within the SW-PBS framework is unknown. Second, additional research is needed on tier 2 intervention efficacy again with an eye toward embedding practices within a full continuum (Bruhn et al. 2013; Mitchell et al. 2012). Third, ongoing research is needed on the systems of support needed to achieve fidelity and maintain implementation of SW-PBS over time (McIntosh et al. 2013). Finally, as the RTI knowledge base continues to expand with respect to both the early intervention and the special education eligibility process, companion work examining the feasibility, including validated prescriptive measures to target tiered supports will be a valuable addition to the knowledge base. For example, the utility of using brief universal social/emotional screening measures to identify at-risk students paired with brief measures of social/emotional ability to guide tier 2 intervention selection prior to problem behaviors becoming chronic and more intense similar to a curriculum-based measure to screen for academic risk and match student to more intensive supports would be an important addition to the SW-PBS knowledge-base (Chafouleas et al. 2013; Kilgus 2013; Kilgus et al. 2013).

Implications for Practice

The clear and widely established impact of SW-PBS universal practices on improved climate, reductions of problem behavior, increases of appropriate social behavior, improved academic performance, reductions of specific behavioral challenges such as bullying, and the improved overall social-emotional well being of students provides a compelling rationale for schools to implement SW-PBS as prevention/early intervention. It is premature to advocate for SW-PBS as part of a social behavioral RTI process to determine student eligibility for special education under IDEA. However, engaging in a process of early intervention and clearly matching behavioral supports to student need and carefully monitoring progress and altering environments to increase the likelihood of success will allow educators to (a) address student need in a more timely systemic manner and (b) provide intervention response data to assist in more comprehensive special education evaluations when appropriate. Using the landmark recommendations of the Peacock Hill Working Group (1991) for implementation of evidence-based practices for at-risk students and those with EBD along with potential points in creating a social behavioral RTI logic for special education eligibility, Table 2

Table 2 Recommended prevention/early intervention and evaluation through response to intervention (RTI) strategies and essential features of school-wide positive behavior support (SW-PBS)

Recommended social behavior evidence-based practices[a]	Response to intervention as best practice and potential evaluation framework	School-wide positive behavior support essential features
Use of systematic, data-based interventions	Consistent implementation of core curriculum. Additional research-validated instruction based on student need	Universal: School team develops social behavior expectations based on presenting problems, explicitly teaches expectations across all staff and settings, and provides corrective and positive-specific feedback across multiple opportunities to practice Tiers 2 and 3: Social skills instruction, self-management, cognitive-behavioral interventions, academic supports, individualized behavior plans based on a functional behavioral assessment
Continuous assessment and monitoring of progress	Quarterly to weekly probes, group data aggregated to examine overall student progress, individual data visually analyzed for trend and progress	Universal: Multiple data sources examined to identify behavioral needs and to monitor progress for all students, settings, and staff Tiers 2 and 3: Data-decision rules established to identify students who require additional supports including screening and teacher referral. Individual student data visually analyzed for trend and progress. Cross student data evaluated to create system efficiencies
Provision for practice of new skills	Core curriculum plus dedicated intervention time for tiers 2 and 3	Universal: Development of a year-long instructional plan where ALL school staff teach, build in practice opportunities across settings, and provide high rates of specific positive praise. Lessons taught within and across all school settings Tiers 2 and 3: All small group and individual strategies connected and aligned to universal expectations and strategies to promote maintenance and generalization
Treatment matched to problem	Benchmark tests linked to level of support, and nonresponders receive more intensive support	Universal social-behavior curriculum annually reviewed and updated based on data, teacher, and student input. Data-based decision-making to match tier II and III supports to function and/or problem type of students
Multicomponent treatment	Core curriculum incorporates effective instruction and scaffolds prior learning. Tiers 2 and 3 provide additional and/or more intensive instructional support	Continuum of supports whereby all students receive universal supports (instruction plus practice with feedback); Tiers 2 and 3 support varied based on student need and guided by assessment. Connections to mental health and other student and family support agencies also included when indicated
Programming for transfer and maintenance	Core curriculum plus additional supports should lead to student fluency within academic skill allowing natural maintenance and generalization opportunities	Linking all tiers 2 and 3 support to universals provides a school-wide environment that incorporates instruction, practice opportunities, and feedback on student use of pro-social skills allowing natural maintenance and generalization opportunities
Commitment to sustained intervention	Overall achievement data guide curriculum and instructional strategies	School and district leadership teams commit to a multiyear process. Team action plans reflect short- and long-term goals
Long term, multilevel approaches to address the issue or problem	Core curriculum plus differentiated linked instruction in place across school year. Students have access to tiers 2 and 3 support at first signs and confirmation of nonresponse	School teams supported by district, region and state SW-PBS supports, school improvement plans include social-behavioral targets, district improvement plans support school social-behavioral targets, linkages to multiagencies along the continuum of supports through logical and strategic connect points established

[a] Peacock Hill Working Group 1991

provides a summary of related SW-PBS practices that allow school teams to build state-of-the art school-wide systems of positive behavior support and increase the likelihood of student success.

While the challenges of students at-risk and those displaying chronic and intense behavioral problems in schools are many, and the outcomes of past, specialized interventions for those students identified with an EBD are bleak, the importance of addressing early patterns of behavioral risk to reduce the unfortunate trajectory of more chronic and intense social/emotional problems is an essential component of education today (Lewis et al. 2010). While typically requiring multiple years to build a complete continuum of behavioral supports through the problem-solving framework of SW-PBS, the documented multiple student and staff benefits certainly justify the investment. As reported above, SW-PBS has documented: (a) improvements in both social and academic behavior among students, (b) reduction in the numbers of students "at-risk," (c) increased staff job satisfaction and (d) overall improvements in school climate. Over time, as educators build a continuum of behavior supports, fewer students display problem behavior within the classroom allowing educators more time to teach and fewer students require more intensive and individualized supports reducing the burden on administrators and specialists (Lewis et al. 2011). While there is no panacea to reduce all behavioral problems and associated risk, adopting a behavioral response to intervention framework, educators can be successful in improving the lives of their students.

Acknowledgment The preparation of this manuscript was supported in part by the Technical Assistance Center on Positive Behavioral Interventions and Supports and a grant from the Office of Special Education Programs, the US Department of Education (H326S980003). Opinions expressed herein do not necessarily reflect the position of the US Department of Education, and such endorsements should not be inferred. For information about the Center, go to www.pbis.org, or for information related to this manuscript, contact Tim Lewis at lewistj@missouri.edu.

References

Algozzine, B., Wang, C., & Violette, A. S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavioral Interventions, 13,* 3–16.

Anderson, C., Lewis-Palmer, T., Todd, A., Horner, R. H., Sugai, G., & Sampson, N. (2011). *Individual student systems evaluation tool*. Educational and Community Supports: University of Oregon.

Barrett, S., Bradshaw, C., & Lewis-Palmer, T. (2008). Maryland state-wide PBIS initiative. *Journal of Positive Behavior Interventions, 10,* 105–114.

Bohanon, H., Fenning, P., Carney, K., Minnis, M., Anderson-Harriss, S., Moroz, K., Hicks, K., Kasper, B., Culos, C., Sailor, W., & Piggott, T. (2006). School-wide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions, 8*(3), 131–145.

Bradley, R., Henderson, K., & Monfore, D. A. (2004). A national perspective on children with emotional disorders. *Behavioral Disorders, 29,* 211–223.

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008a). The impact of school-wide Positive Behavioral Interventions and Supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23*(4), 462–473.

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008b). Implementation of school-wide Positive Behavioral Interventions and Supports (PBIS) in elementary schools: Observations from a randomized trial. *Education & Treatment of Children, 31,* 1–26.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide positive behavioral interventions and supports: Findings from a group-randomized effectiveness trial. *Prevention Science, 10*(2), 100–115

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of school-wide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12,* 133–148.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (in press). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*. doi:10.1542/peds.2012-0243.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130(5),* 1136–1145.

Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2013). A review of tier 2 interventions conducted within multi-tiered models of prevention evidencing a primary behavioral plan. *Journal of Emotional and Behavioral Disorders*. doi:10.1177/1063426613476092. (Advance online publication)

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J. (2013). Direct behavior rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology, 51,* 367–385.

Chapman, D., & Hofweber, C. (2000). Effective behavior support in British Columbia. *Journal of Positive Behavior Interventions, 2*(4), 235–237.

Cheney, D., Flower, A., & Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *The Journal of Special Education, 42(2),* 108–126.

Colvin, G., Kameenui, E. J., & Sugai, G. (1993). School-wide and classroom management: Reconceptualizing the integration and management of students with behavior problems in general education. *Education and Treatment of Children, 16,* 361–381

Curtis, R., Van Horne, J., Robertson, P., & Karvonen, M. (2010). Outcomes of a school-wide positive behavior support program. *Professional School Counseling, 13*(3), 159–164.

De Pry, R. L., & Sugai, G. (2002). The effect of active supervision and precorrection on minor behavioral incidents in a sixth grade general education classroom. *Journal of Behavioral Education, 11,* 255–267.

Duda, M. A., Dunlap, G., Fox, L., Lentini, R., & Clarke, S. (2004). An experimental evaluation of positive behavior support in a community preschool program. *Topics in Early Childhood Special Education, 24*(3), 143–155.

Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M (2007). Response to intervention: Examining classroom behavior support in second grade. *Exceptional Children, 73*(3), 288–310.

Farkas, M., Simonson, B., Migdole, S., Donovan, M., Clemens, K., & Cicchese, V. (2012). Schoolwide positive behavior support in an alternative school setting: An evaluation of fidelity, outcomes and social validity of Tier I implementation. *Journal of Emotional and Behavioral Disorders, 20*(4), 275–288.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Forness, S., & Knitzer, J. (1992). A new proposed definition and terminology to replace "serious emotional disturbance". Individuals with disabilities education act. *School Psychology Review, 21,* 12–20.

Gilliam, W., & Shabar, G. (2006). Preschool and child care expulsion and suspension rates and predictors in one state. *Infants and Young Children, 19,* 228–245.

Goodman, S. (2013, June). *District-level adoption: The engine to drive sustainable change*. Presentation at the annual Missouri school-wide positive behavior support conference, Lake of the Ozarks, MO.

Gresham, F. M. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children, 28,* 328–344.

Gresham, F. M. (2007). Response to intervention and emotional and behavioral disorders: Best practices in assessment for intervention. *Assessment for Effective Intervention, 32,* 214–222.

Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks: Corwin.

Hawken, L. S., Vincent, C. G., & Schumann, J. (2008). Response to intervention for social behavior: Challenges and opportunities. *Journal of Emotional and Behavioral Disorders, 16*(4), 213–225.

Hirsch, E. J., Lewis-Palmer, T., Sugai, G., & Schnacker, L. (2004). Using school bus discipline referral data in decision making: Two case studies. *Preventing School Failure, 48*(4), 4–9.

Horner, R. H., & Sugai, G., (2005). School-wide positive behavior support: An alternative approach to discipline in schools. In L. Bambara & L. Kern (Eds.), *Positive behavior support* (pp. 359–390). New York: Guilford Press.

Horner, R. H., Todd, A., Lewis-Palmer, T., Irvin, L., Sugai, G., & Boland, J. (2004). The school-wide evaluation tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Intervention, 6*(1) 3–12.

Horner, R., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A., & Esperanza, J., (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11,* 133–145.

Kauffman, J. M., Simpson, R. L., & Mock, D. (2009). Problems related to underservice: Rejoinder. *Behavioral Disorders, 34*(3), 172–180.

Kilgus, S. P. (2013). Use of direct behavior ratings as the foundation of tier 2 service delivery. *Journal of Curriculum and Instruction, 7,* 79–99.

Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial validation of the Social and Academic Behavior Risk Screener for elementary grades. *School Psychology Quarterly, 28,* 210–226.

Lane, K. L., Carter, E. W., Pierson, M. R., & Glaeser, B. C. (2006). Academic, social, and behavioral characteristics of high school students with emotional disturbances or learning disabilities. *Journal of Emotional and Behavioral Disorders, 14*(2), 108–117.

Lewis, T. J. (2010, March). *Are we there yet? Mapping the SW-PBS course for the long haul*. Invited presentation at the 7th annual association for positive behavior support conference. St. Louis, MO.

Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive schoolwide management. *Focus on Exceptional Children, 31*(6), 1–17.

Lewis, T. J., Colvin, G., & Sugai, G. (2000). The effects of precorrection and active supervision on the recess behavior of elementary school students. *Education and Treatment of Children, 23,* 109–121.

Lewis, T. J., Powers, L. J., Kelk, M. J., & Newcomer, L. (2002). Reducing problem behaviors on the playground: An investigation of the application of school-wide positive behavior supports. *Psychology in the Schools, 39,* 181–190.

Lewis, T. J., Jones, S. E. L., Horner, R. H., & Sugai, G. (2010). School-wide positive behavior support and students with emotional/behavioral disorders: Implications for prevention, identification and intervention. *Exceptionality, 18*(2), 82–93.

Lewis, T. J., Bradley, L., & Freeman, C. (2011, Oct). *Using data to design and monitor tier 2 supports: From classroom to small group*. Presentation at the eighth annual school-wide PBS implementation forum. Chicago, IL.

Lohrmann-O'Rourke, S., Knoster, T., Sabatine, K., Smith, D., Horvath, G., & Llewellyn, G. (2000). School-wide Application of PBS in the Bangor Area School District. *Journal of Positive Behavior Interventions, 2*(4), 238–240.

Luiselli, J. K, Putnam, R. F, Handler, M. W, & Feinberg, A. B. (2005). Whole-school positive behaviour support: Effects on student discipline problems and academic performance. *Educational Psychology, 25*(2–3), 183–198.

Maag, J. W., & Katsiyannis, A. (2008). The medical model to block eligibility for students with EBD: A response-to-intervention alternative. *Behavioral Disorders, 33*(3), 184–194.

Mathur, S. R. (2007). Understanding emotional and behavioral disorders: Are we paying the cost of borderline ethics? *Education and Treatment of Children, 30,* 11–26.

McIntosh, K., Chard, D. J., Boland, J. B., & Horner, R. H. (2006). Demonstration of combined efforts in school-wide academic and behavioral systems and incidence of reading and behavior challenges in early elementary grades. *Journal of Positive Behavioral Interventions, 8,* 146–154.

McIntosh, K., Flannery, K. B., Sugai, G., Braun, D., & Cochrane, K. L. (2008a). Relationships between academics and problem behavior in the transition from middle school to high school. *Journal of Positive Behavior Interventions, 10*(4), 243–255.

McIntosh, K., Horner, R. H., Chard, D. J., Dickey, C. R., & Braun, D. H. (2008b). Reading skills and function of problem behavior in typical school settings. *Journal of Special Education, 42,* 131–147.

McIntosh, K., Sadler, C., & Brown, J. A. (2012). Kindergarten reading skill level and change in as risk factors for chronic problem behavior. *Journal of Positive Behavior Interventions, 14*(1), 17–28.

McIntosh, K., Mercer, S. H., Hume, A. E., Frank, J. L., Turri, M. G., & Mathews, S. (2013). Factors related to the sustained implementation of schoolwide positive behavior support. *Exceptional Children, 79,* 293–311.

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., et al. (2010). Lifetime prevalence of mental disorders in U.S. Adolescents: Results from the national comorbidity survey replication—adolescent supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(10), 980–989.

Miller-Richter, M., Lewis, T. J., & Hagar, J. (2012). The relationship between principal leadership skills and school-wide positive behavior support: An exploratory study. *Journal of Positive Behavior Interventions, 14,* 69–77.

Mitchell, B. S., Stormont, M. A., & Gage, N. A. (2011). Tier 2 interventions within the context of school-wide positive behavior support. *Behavioral Disorders, 36*(4), 241–261.

National Research Council and Institute of Medicine. (2009). Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities. Committee on the prevention of mental disorders and substance abuse among children, youth and young adults: Research advances and promising interventions. In M. E. O'Connell, T. Boat, & K. E. Warner (Eds.), *Board on children, youth and families, division of behavioral and social sciences and education*. Washington, DC: The National Academies Press.

Nelson, J. R., Martella, R., & Galand, B. (1998). The effects of teaching school expectations and establishing a consistent consequence on formal office disciplinary actions. *Journal of Emotional and Behavioral Disorders, 6,* 153–161.

Peacock Hill Working Group. (1991). Problems and promises in special education and related services for children and youth with emotional or behavioral disorders. *Behavioral Disorders, 16,* 299–313.

Putnam, R. F., Luiselli, J. K., & Sunderland, M. (2002). Longitudinal evaluation of behavior support intervention in a public middle school. *Journal of Positive Behavior Interventions, 4,* 182–188.

Putnam, R. F., Handler, M. W., Ramirez-Platt, C. M., & Luiselli, J. K. (2003). Improving student bus-riding behavior through a whole-school intervention. *Journal of Applied Behavior Analysis, 36,* 583–589.

Robers, S., Kemp, J., & Truman, J. (2013). *Indicators of school crime and safety: 201*2 (NCES 2013-036/NCJ 241446). National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Rose, L. C., & Gallup, A.M. (2007). The 39th annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 89*, 33–48.

Simonson, B., Britton, L., & Young D. (2010). School-wide positive behavior support in an alternative school setting: A case study. *Journal of Positive Behavior Interventions, 12,* 180–191.

Stichter, J. P., Lewis, T. L., Richter, M., & Johnson, N. (2006). Assessing antecedent variables: The effects of instructional variables on student outcomes through in-service and peer coaching professional development models. *Education and Treatment of Children, 29,* 665–692.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A., Turnbull, H. R., Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2,* 131–143.

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of School-wide Positive Behavioral Interventions and Supports (SWPBIS) on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics and Adolescent Medicine, 116*(2), 149–156.

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavior Disorders, 13*(2), 79–96.

Walker, H. M., & Bullis, M. (1991). Behavior disorders and the social context of regular class integration: A conceptual dilemma? In J. Lloyd, N. Singh, & A. Repp (Eds.), *The regular education initiative: Alternative perspectives on concepts, issues and models* (pp. 75–94). Champaign-Urbana: Sycamore.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4,* 193–256.

Walker, H. M., Ramsey, E., & Gresham, R. M. (2004). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont: Wadsworth.

Toward a Unified Response-to-Intervention Model: Multi-Tiered Systems of Support

Matthew K. Burns, Shane R. Jimerson, Amanda M. VanDerHeyden and Stanley L. Deno

As discussed in Chap. 1, response to intervention (RTI) has become a national movement. Many elementary schools in the country are implementing or attempting to implement an RTI model. A recent national survey of K-12 administrators indicated that approximately two thirds of the schools that completed the survey were either in full implementation or in the process of implementing an RTI model, which was up from 24% in 2007 (Spectrum K-12 Solutions 2010). However, when asked if they had a model from which to work, less than half indicated that they did. This is alarming because how could it be that more schools are implementing RTI than have a plan to do so? From what framework are schools implementing their RTI models?

The authors' experience suggests that schools continue to create fundamental difficulties when implementing RTI. The final chapter in the first *Handbook of Response to Intervention* in 2007 (Burns et al.) articulated a unified model for RTI that was based on the IDEAL model of problem-solving in which school personnel (a) identify the problem, (b) define the problem, (c) explore alternative solutions to the problem, (d) apply a solution, and (e) *l*ook at the effects of the application (IDEAL; Bransford and Stein 1984). The authors continue to emphasize that RTI is an approach to solve problems using data, but the problems may be at the system, classroom, small-group, or individual level. To be successful, school personnel need to address the systems issues of the RTI framework.

One common difficulty generally observed in K-12 schools is that for many of them, their first step in starting an RTI model is to start a problem-solving team (PST) because they recognize that RTI is a problem-solving process. In authors' opinion, and experience, this is a mistake for two important reasons.

The first reason why it is a mistake to implement a PST as the first step in implementing an RTI model is that effective PSTs are resource intensive. Most PSTs require five to eight professionals to meet for 15–30 min to discuss one student. Consider a typical elementary school with 600 students in K-5. On average, 20% of the students need reading support beyond quality core instruction (Burns et al. 2005), which would equal 120 students. Even a small school of half that size (i.e., 300 students) there would still be 60 students who require additional support.

M. K. Burns (✉)
Department of Educational Psychology, University of Missouri, MO, USA
e-mail: burnsmk@missouri.edu

S. R. Jimerson
Department of Counseling, Clinical, and School Psychology, University of California, Santa Barbara, CA, USA
e-mail: jimerson@education.ucsb.edu

A. M. VanDerHeyden
Education Research & Consulting, Inc, Fairhope, AL, USA
e-mail: amandavande@gmail.com

S. L. Deno
Department of Educational Psychology, University of Minnesota, Minnesota, USA
e-mail: denox001@umn.edu

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3_41

Moreover, many schools have more than 20% who need additional support beyond core instruction. Thus, the challenges presented in trying to schedule a PST meeting for 60–120 students can be recognized. If the team meets once each week and discusses two students each time, that is 45 weeks, which exceeds the length of the school year for most schools. Most schools simply do not have the resources needed to conduct a PST meeting for 60–120 students. Within this context, deploying, supporting, and sustaining interventions for so many children is likely to present numerous challenges and the likelihood of intervention success is remote.

Attempting to treat system problems as individual problems is a costly and common pitfall of RTI implementation for many systems (VanDerHeyden and Tilly 2010). Schools should first rule out the need for grade-wide or classwide intervention first and then implement an effective tier 2 intervention that focuses on approximately 15% of the students and targets the intervention with low-level analysis. Having a strong tier 2 will reduce the number from 60 or 120 to 15 to 30, which is a more reasonable number to address. Without an effective tier 2, you cannot have an effective PST process (tier 3).

The second reason why it is not a good idea to start an RTI initiative by starting a PST is that it ignores potential systemic learning problems. In many of the schools in which the authors have worked, more than 20% of the students needed additional support. In fact, many schools have 50% or more of students needing supplemental instruction to master core instructional objectives. Consider a classroom of 25 students in which 15 students are identified as needing intervention. Could school personnel pull 15 of 25 students to run a small-group intervention? Of course not. In this common example, there is a system issue and that system issue must be addressed before starting tier 2 or tier 3 interventions. Thus, an RTI model should begin by focusing on core instruction. As stated above, there cannot be an effective tier 3 intervention without an effective tier 2 intervention system, but an effective tier 2 is dependent on an effective tier 1 (core instruction). Without good core instruction, nothing else matters! The importance of quality core instruction is often missed when RTI initiatives begin by starting a PST.

Multi-Tiered System of Support

Perhaps school personnel should conceptualize RTI differently than many have before. RTI came out of special education legislation that stated that a local educational agency may "use a process based on the child's response to scientific, research-based intervention" (34 C.F.R. § 300.307[a][2]) as part of the learning disability (LD) eligibility process. RTI is the process of providing quality instruction, implementing interventions matched to student need, and using student response data to make instructional and important educational decision (Batsche et al. 2005). Many states are adopting the term multi-tiered system of support (MTSS) because it focuses on providing instruction and intervention, rather than using data to identity disabilities.

RTI and MTSS are not just different names for the same thing. RTI came out of special education law, but MTSS focuses on general education. RTI is assessment oriented, but MTSS emphasizes providing services. Both rely on screening all students, providing tiered interventions, monitoring student progress, and using a problem-solving framework, but MTSS has a more explicit focus on general education. In 2006, RTI is defined as the systematic use of assessment data to enhance learning for all students (Burns and VanDerHeyden 2006). That definition of RTI focused on prevention of academic failure through effective instruction for all students, and was consistent with the contemporary conceptualization of MTSS, which is the term and concept that are now preferred. Educational issues that gave rise to RTI and now suggest the need for MTSS are discussed below.

LD Identification

Education today is different than it was in 2007 when the first edition of this book appeared. First, LD identification has fundamentally changed. The discrepancy model that was first adopted in federal special education regulations in 1977 as a result of a political compromise (Gresham et al. 2005) has been discredited (Aaron 1997; Algozzine and Ysseldyke 1982, Fletcher et al. 1998). The concept of identifying LD by measuring student learning was not new in 2007 (e.g., Fuchs and Fuchs 1998; Vellutino et al. 1996), but it remains perplexingly controversial. However, even those who are fundamentally opposed to using data compiled through RTI processes to identify LD accept that student response to research-based interventions should be part of the identification process (e.g., Kavale et al. 2008). Thus, RTI methods will likely remain at least a part of LD identification evaluations for the foreseeable future and it is important to focus on student outcomes.

Accountability and Outcomes

It is not hyperbolic to state that the *Nation at Risk* (US Department of Education, 1983) report has forever changed education in this country by creating a call for outcomes-based reform. Moreover, there has been a general increased interest in accountability of government agencies, including education, over the past 20 years (McDermott 2011). Salvia and Yssledyke (2001) defined accountability as documentation for people in authority that "desired goals are being met" (p. 644). Desired goals were not specified for children with disabilities before 1997 amendments to IDEA, but that changed when it was mandated that children who participated in special education also participate in state accountability assessments.

Although RTI was articulated in special education law (IDEA 2004), it was conceived in the *No Child Left Behind Act* (NCLB 2001) and the many policy documents surrounding and informing these two key pieces of legislation (Burns and Gibbons 2012). NCLB specified that the achievement of all children be measured including those with disabilities, those who are English language learners, and those from low socioeconomic backgrounds. The emphasis on measuring student achievement and progress was consistent with the data-based decision-making movement that began in the 1970s, and was endorsed by the Presidents Commission on Excellence in Special Education (PCESE, 2002). IDEA (2004) specified that children could be diagnosed as LD if "the child fails to achieve a rate of learning to make sufficient progress to meet state-approved results in one or more of the areas identified in paragraph (a)(1) of this section when assessed with a response to scientific, research-based intervention process" (P.L. 108–446, §§ 300.309), which directly linked IDEA to NCLB and RTI to accountability. In fact, the language in all of these documents has been characterized as "mutually referential" (Kovaleski et al. 2013, p.???). The focus on increasing student outcomes has further evolved RTI as a method of system improvement and prevention that schools are beginning to adopt and that is endorsed here.

Implementing MTSS

MTSS models are driven by effective grade-level teams functioning as professional learning communities (PLCs) that examine student outcome data within a culture of collaboration to enhance student learning (DuFour 2005). PLCs can examine core instructional issues, but PSTs generally focus on outcomes for individual students, which puts core instruction directly on the assessment plan of an MTSS. Moreover, PLCs are responsible for examining student data within an MTSS to determine who needs tier 2 or tier 3 intervention, what type of intervention a student needs, and if the student is making sufficient progress. If classroom teachers are more closely involved in designing interventions within MTSS, then there would be a more direct link between instruction and intervention, which was often missing in previous intervention models.

Sugai and Horner (2009) outline the essential attributes of an MTSS. The focus of their work is positive behavior support, but the framework applies to any MTSS model and is outlined in Fig. 1. As can be seen in the figure, MTSS relies

Defining features of Multi-Tiered System of Supports (MTSS)

1. Interventions supported by scientifically-based research.
2. Interventions organized along a tiered continuum that increases in intensity (e.g., frequency, duration, individualization, specialized supports, etc.)
3. Standardized problem-solving protocol for assessment and instructional decision making.
4. Explicit data-based decision rules for assessing student progress and making instructional and intervention adjustments.
5. Emphasis on assessing and ensuring implementation integrity.
6. Regular and systematic screening for early identification of students whose performance is not responsive to instruction.

Fig. 1 Defining features of multi-tiered system of supports (MTSS). (Based on Sugai and Horner 2009)

on several aspects of quality instruction, but also involves implementing a standardized problem-solving framework. Several generic problem-solving models exist in the literature, but perhaps the most specific details the steps as: (a) identify the problem, (b) define the problem, (c) explore alternative solutions to the problem, (d) apply a solution, and (e) look at the effects of the application (IDEAL; Bransford and Stein 1984). When this model is applied to solving the problems addressed through MTSS, the problem to be solved is how "to eliminate the difference between 'what is' and 'what should be' with respect to student development" (p. 38, Deno 2002), and MTSS is an attempt to identify resources necessary for sufficient student learning to occur. This chapter examines current practice and research using the steps in the IDEAL model and uses this as the basis for recommending MTSS practices that would construct a unified model.

Identify the Problem

Current Practice Student outcome data are becoming more prominent in designing interventions for individual students and groups of children (Shapiro 2000), and are less likely based on high inferences such as measures of cognitive processes but are more direct assessments of the academic problem (GET). There are many depictions of MTSS models in the literature that emphasize data collection as a means to identify problems for individual students (Howe, Scierka et al. 2003; Ikeda et al. 1996; Marston et al. 2003), and most districts have implemented universal screenings for reading and mathematics, and some for behavior (Kettler et al. 2014). Thus, assessment appears to play a prominent role in districts in which MTSS practices are engaged, and is becoming commonplace in K-12 schools.

Research-Based Practices General outcome measures or curriculum-based measurement (CBM) are especially useful in identifying areas of skill deficits for children. CBM could serve as an effective first step in any problem-solving effort in that data could be compared to various standards and be used to identify individual children in need of additional intervention (Shinn 2008). Moreover, CBM has been identified as an essential component of any effective MTSS model (Burns and Ysseldyke 2005).

The research regarding CBM is valuable from both a psychometric and instructional perspective. Data obtained from CBM have been shown to be sufficiently reliable for instructional decisions among various student populations (Deno 2005), and using those data for instructional decisions led to increases in student learning (Fuchs and Fuchs 1986). Data obtained from brief assessments of academic skills for all students can be used to identify children with potential difficulties. Thus, the first step of the problem-solving process is to screen the academic skills of all students, called universal screening. Research has supported effectiveness of using screening data in enhancing the learning outcomes of all children (Ardoin et al. 2005, VanDerHeyden and Burns 2005; VanDerHeyden et al. 2003).

There are a variety of CBM packages, many of which are available for free: Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Kaminski and Good 2002—https://dibels.uoregon.edu/), AIMSweb (2010—http://www.aimsweb.com), Easy CBM (Alonzo et al. 2006—http://www.easycbm.com/), and Computer-Based Assessment System for Reading (Kroll et al. 2006—www.faip.umn.edu). A thorough review of various screening measures can be found at the National Center on Response to Intervention (NCRTI) website, http://rti4success.org/screeningTools.

As stated above, data need to be collected for all children on a continuous basis in order to screen for academic deficits and CBM is an effective approach for continuous progress monitoring. However, other assessments could also serve as screening instruments including Measures of Academic Progress for Reading (MAP-R, Northwest Evaluation Association, 2003), STAR-Reading (Renaissance Learning, 2003), STAR-Math (Renaissance Learning, 2005), the Iowa Test of Basic Skills (ITBS; Hieronymus et al. 1979), or the Gates–MacGinitie Reading Test (GMRT; MacGinitie et al. 2000. One cannot go into detail about the strengths and relative shortcomings or each, but the tools listed above were highly rated by the NCRTI. Research is needed to identify which approach is superior, but likely advantages of the former over the latter is the sensitivity to growth, ease of use, lower costs, dynamic nature, and the ability to inform other aspects of problem-solving (Shinn 2008).

Define the Problem

Current Practice Clearly and explicitly defining the problem is the "key to success" of problem-solving (Deno 2002, p. 46). Although many MTSS models articulated in the literature involve what is called problem-solving, few convey steps to defining the problem. The Screening to Enhance Educational Progress (STEEP; VanDerHeyden et al. 2003) begins this process by first examining if the difficulty is specific to the child or the classroom of children and then determining if the deficit is primarily due to a lack of skill or lack of motivation (Ardoin et al. 2005; VanDerHeyden et al. 2003; Witt 2002).

Research-Based Practices The model proposed by Shapiro (2010) suggests that assessment data are gathered to evaluate the academic environment, instructional placement, and instructional modifications to define the problem. Many elementary schools screen their students' academic skills, but the data are often not compared to meaningful criteria and may not be used to their full potential (Glover and Albers 2007). PLCs, which are a critical component of an effective MTSS, provide a forum for educators to collectively problem solve at the school, classroom, and student level, and should review universal screening data to do so. PLCs focus on student outcome data and create a culture of collaboration to enhance student learning (DuFour 2005). However, most PLCs struggle to identify common assessments, criteria with which to judge student proficiency, and a process to collaboratively analyze data and improve student learning (DuFour et al. 2005; Love 2009).

Defining the problem is essentially analyzing the problem with sufficient depth. Thus, PLCs should analyze student problems by answering the following three questions as outlined by Burns and Gibbons (2012):

1. Is there a class-wide need?
2. Among the students who need an intervention, what is the category of the problem?
3. What is the causal variable?

Class-Wide Need The first step in analyzing the problem within an MTSS is to determine if there are class-wide needs. VanDerHeyden and colleagues (VanDerHeyden and Burns 2005, 2010; VanDerHeyden et al. 2003) defined a class-wide need as individual student difficulties being the result of a potential systems issue rather than individual students with low capabilities, and operationally defined it as a class median that is below a given standard. Previous research has consistently found that implementing a brief (e.g., 15–20 min) class-wide intervention led to substantial gains in student growth and decreases in the number of students requiring intervention (Burns et al. in press; VanDerHeyden and Burns 2005; VanDerHeyden et al. 2012; VanDerHeyden et al. 2003).

Category of the Problem Tier 2 interventions generally target the same skills that are taught during core instruction and should also be aligned with the benchmarking measures used for that grade. However, tier 2 interventions should target critical basic reading skills, such as those identified by the NRP (NICHD 2000), phonemic awareness, phonics, reading fluency, comprehension, and vocabulary, or discrete mathematics objectives (VanDerHeyden and Burns 2010). Therefore, PLCs examine data from the five NRP areas and from measures of specific mathematics objectives to determine the intervention target. Targeting the intervention led to increases in student skill in reading (Chap. 17) and mathematics (VanDerHeyden and Burns 2005), and recent research found larger effects for targeted reading interventions than for evidence-based interventions that were more comprehensive in nature (Hall and Burns 2014). Howell and Nolet (1999) articulated an extensive instructional decision-making model called curriculum-based evaluation (CBE) that emphasizes task analyses, direct observation, and systematic hypothesis testing. Although research supports the components of this approach, few studies have examined outcomes associated with the model in its entirety.

Causal Variable No single approach intervention, no matter how well researched, easy to implement, or intensive, is effective for all students (Fuchs and Fuchs 2005). There will always be a group of students who receive a tier 2 intervention for whom the intervention is not effective. School-based personnel have to realize that tier 2 is not designed to address the needs of all students, but it is design to address the needs of most students who are experiencing difficulties. Brief experimental analysis (BEA) is a useful tool with which in-depth problem analysis can be conducted for tier 3 interventions in order to determine the causal variable. The term causal variable means the environmental or instructional variable that is most closely related to the problem (Burns and Gibbons 2012). BEA is a brief functional analysis approach with which hypotheses regarding academic performance deficits can be tested to isolate and manipulate specific mechanisms related to academic performance. For example, Daly et al. (1997) advocated using BEA for functional analysis of academic performance deficits such as lack of motivation, insufficient practice, inadequate feedback, novel instructional materials, and difficult instructional materials. BEA has been shown to be effective in identifying appropriate reading interventions for kindergarteners (Petursdottir et al. 2009), oral reading fluency interventions for students in the primary grades (Daly et al. 1999, 2006), as well as letter formation interventions for second graders (Burns et al. 2009).

Explore Alternative Solutions to the Problem

Current Practice Almost all MTSS models that currently exist in K-12 schools use a multidisciplinary PST to generate alternatives solutions for student problems (Burns and Ysseldyke 2005). Most publications describing various PST approaches specifically discuss who should be members of the team, but few discuss what specific process the team uses to generate and explore potential solutions. Those articles that do address a problem-solving process tend to use vague language such as after identifying the problem, "the next step is to identify why it is occurring. For problems at a low level of intensity, hypotheses about why the problem is happening may be derived informally. As problems become more intense, a more rigorous and systematic problem analysis procedure will be necessary" (Heartland Area Education Agency 11, 2006, p. 101). Moreover, there seems to be a wide range of activities in which educators engage in the name of problem-solving (Burns et al. 2005).

Research-Based Practices Research has consistently supported the use of PSTs to generate potential solutions (Burns and Symington 2002; Ikeda and Gustafson 2002; Marston et al. 2003), but an empirical investigation as to whether PSTs are a critical component of an MTSS has yet to be completed. As stated above, PLCs should be examining student data at tiers 1 and 2 in order to (a) identify class-wide needs, (b) determine who needs additional support beyond core instruction, (c) identify appropriate interventions for small-groups of students, and (d) determine which students receiving a tier 2 intervention are successful and which need even more intensive support. Once a PLC determines that additional support is needed (i.e., tier 3), then a referral to a PST can occur.

Apply a Solution to the Problem

Current Practice Once interventions are found to be effective for an individual student, they are implemented over an extended period, but the delivery system can vary substantially between models. Some models match delivery system with student need and may include special education services as a delivery option (Lau et al. 2006; Tilly 2002); and others more or less restrict remedial efforts to general education but could utilize individual, small-group, or class-wide interventions (Kovaleski et al. 1996). Generally speaking, interventions implemented to solve a problem are categorized as problem-solving or standard protocol, with the defining difference being the uniformity of remedial efforts (Fuchs et al. 2003). Although this dichotomy is probably artificial (Christ et al. 2006), many MTSS models currently in place probably fit into one or the other category.

Research-Based Practices Research has consistently supported the effectiveness of both the problem-solving and standard-protocol approaches to MTSS (Burns et al. 2005), which makes the conversation about selecting one over the other somewhat moot. Thus, the efficiency, rather than effectiveness, should probably drive the delivery of interventions within an MTSS. Class-wide interventions (Chap. 15) and small-group tier 2 interventions (Chaps. 18 and 19) often rely on standardized protocols for intervention and tier 3 should be highly individualized.

Look at the Effects of the Application

Current Practice The culture of schools in this country has increasingly embraced data-based decision-making (Ysseldyke et al. 2006), but many of those decisions are based on state-mandated group achievement tests administered for accountability purposes. More instructionally relevant measures such as CBM have become more common, probably due to the readily available and easily used electronic data warehouses for PK-12 schools (Ysseldyke and Mcleod 2007). Moreover, some districts have developed somewhat sophisticated analyses to link CBM data to state accountability test scores and to derive benchmark criteria (Bollman et al. 2007).

Research-Based Practices The recent use of CBM for data-based decision-making follows a long line of research supporting its effectiveness (Deno 2005; Fuchs and Fuchs 1986). Recent efforts have demonstrated a moderate to strong link between CBM scores and state accountability test scores (Stage and Jacobsen 2001; McGlinchty and Hixson 2003), but standards for educational assessment suggest the need to directly examine the accountability and eligibility decision-making utility of CBM (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 1999). Thus, additional research is needed. However, research has consistently supported that CBM is ideally suited to measure the effectiveness of interventions within a problem-solving model.

Recent developments in psychometric aspects of CBM shed some light on current practices. Perhaps most importantly is the effect that standard error of measure has on CBM scores. In typical assessment situations, approximately 8 weeks' worth of data are needed before the value of the standard error of measure of student growth rates is smaller than the value of the slope of growth (Christ 2006). In other words, a slope of 1.5 words per minute per week, which indicates the child increases her or his reading fluency by 1.5 words per minute each week, would likely have a true score range of −0.5 to 3.5 or larger until data are collected for 8 weeks.

Unified MTSS Model: Problem-Solving

Implementation integrity remains a substantial threat to MTSS implementation unless the field can articulate a common model, or at least the core components of an effective practice. As a result of a review of the research literature and current practice a three-tiered model that infuses the principles of problem-solving throughout is endorsed. The specific activities associated with different problem-solving components are outlined in Fig. 2.

Matrix Representing a Multi-Tiered System of Support and the IDEAL Problem Solving Model (Bransford & Stein, 1984)

	IDEAL Problem Solving Model				
	Identify the Problem	Define the problem	Explore alternative solutions	Apply a solution	Look at the effects of the application
Tier 1	Implement quality core instruction and conduct universal screenings	Collect data to rule out class-wide or curricular problems	Generate potential class-wide interventions if necessary	Implement class-wide remedial interventions or make instructional modifications	Continue benchmark assessment to determine if the class progresses
Tier 2	Examine universal screening data to identify struggling students	Collect data regarding student proficiency categories of specific skills areas (e.g., decoding, reading fluency, and single-digit multiplication)	Generate a list of evidence-based strategies to intervene for a small group of students	Implement explicit instructional strategies to address the problem area for a small-group of children	Outcome assessment to examine progress at least every other week
Tier 3	Examine progress monitoring data to determine if students are making sufficient progress	Conduct brief experimental analyses to determine the instructional variable that most closely related to the problem	Generate a list of evidence-based individualized interventions	Implement evidence-based intensive individualized interventions to address the problem area	Frequent (twice weekly) outcome assessment to examine progress

Fig. 2 Matrix representing a multi-tiered system of support and the IDEAL problem-solving model. (Bransford and Stein 1984)

Tier 1

The first tier of an MTSS model should be defined by quality core curriculum and universal screening of all children with instructionally sensitive and psychometrical adequate tools. Fortunately, an extensive literature exists regarding both, so readers are referred to other sources for specific information. The primary aspect of problem-solving that occurs within the first tier is *identifying a problem*. Student skills should be compared to benchmark criteria to determine if a problem exists, but identifying the problem and searching for a solution occurs in later stages of problem-solving and subsequent tiers in MTSS. The previous statement is only true; however, if the problem is determined to be specific to the individual child. Examining if the deficit is class wide would rule out instructional and curricular explanations for the individual child, but could suggest the need for an intervention or modification in tier 1 that would affect all children in the class.

Tier 2

Define the Problem On average, approximately 20 % of children will not respond adequately to tier I instruction (Burns et al. 2005) and should receive a tier II intervention. After identifying the problem, further defining of the difficulty should occur by first analyzing the magnitude of the discrepancy between what is expected and how the child is performing. Next, data for the child's classroom should be examined to determine if the low performance is specific to the child or a residual effect of an ineffective tier I. This is done by comparing the mean of the classroom benchmark data to a criterion to assure that it is the child and not the class that lacks proficiency. Other data should be collected such as curriculum-based assessment (Gickling and Havertape 1981), criterion-referenced assessment, and an ecological analysis of contextual influences to better understand and define the problem area for individual children.

Explore Alternatives If the class median falls below the benchmark criterion, then the intervention should be delivered to the entire class. If the median score is above the benchmark criterion, then the intervention should be delivered to the child. Potential interventions should be examined within the framework of the National Reading Panel (2000) by assessing the child's skills in phonemic awareness, phonics, fluency, vocabulary, and reading comprehension. After identifying which area is most likely linked to the reading deficit, a specific intervention can be attempted.

Apply a Solution Interventions within tier 2 should be delivered in a small-group format based on NRP areas for reading and specific objectives for mathematics. Generally speaking, these groups will have three to four members, but should be limited to six members. After grouping children according to needs, explicit instruction in the deficit area should occur for at least 30 to 60 min each day with at least supervision by a highly qualified teacher.

Look at Effectiveness Outcome assessment within tier 2 needs to occur at least every other week and should address the same general outcome measure as used in tier 1 universal screenings. The reason that consistency between tiers is important is because the level of student skill and slope of growth (dual discrepancy) should both be compared to the general population, which requires that the data obtained within that general population be directly comparable to those used to monitor progress in tier 2. Those students found to sufficiently respond would either return to tier I or continue with tier 2 support. Those whose skill level falls below a criterion, and whose rate of growth falls below the normative standard would next receive a tier 3 intervention to continue exploring alternative interventions.

Tier 3

Define the Problem Shapiro's (2010) assessment-to-intervention model proposes that curriculum-based assessment data are needed to modify instruction. These data seem critical in tier 3, but other data will be needed as well and could include norm-referenced measures of word reading, reading comprehension, or phonological processing. Again, the severity of the problem can be defined normatively, but baseline and functional data are also required.

Explore Alternatives Interventions in tier 2 are designed to be efficient, in that they focus on standardized approaches for groups of children. Interventions in tier 3 are more clearly focused on effectiveness rather than efficiency because intensive individualized interventions will be explored until some individual or combination of interventions leads to student success. Perhaps the best method to explore interventions over the short term, to then implement over the long term, is BEA. Children who are not successful in tier 2 should be presented to a PST, which adheres to the principles of an effective PST (Burns et al. 2014), after a BEA has identified potential interventions. The PST would then brainstorm how to make those interventions as practical and effective as possible. This approach would match the functional analysis aspect of most problem-solving models.

Apply a Solution Interventions within tier 3 should be limited to small groups of three or less, and could even be delivered in a one-on-one format. Although the exact intervention will likely vary from child to child, it should target the deficit area, explicitly teach the skill, provide frequent opportunities to respond, use materials that provide an appropriate level of challenge, and contain sufficient feedback to inform the child of successes and errors. There is no adequate data source from which to specify the dosage that is appropriate for tier 3, but some policy documents have suggested that tier 3 intervention should consist of at least 30 min of daily instruction beyond the general education core curriculum (Gersten et al. 2008).

Special education services could be utilized in tier 3, but only if they are needed to assure student success. The goal of MTSS is to keep searching until the solution to the child's problem is found, and then to implement the intervention as efficiently as possible. For some children, the level of need and/or intensity of intervention may be such that the child cannot be successful unless special education resources are allocated. At that point, the child would be identified as having a special education disability and special education would be invoked. Kovaleski and colleagues (2013) provide detailed guidelines on how to use RTI data to determine eligibility for LD and to guide individualized educational program development.

Look at Effectiveness CBM data are needed in tier 3. Student progress data should be collected at least weekly and progress toward a goal should be closely monitored. A goal can be established for an individual child based on normative criteria (e.g., average reading rate for children in his or her grade) or criterion-related data (e.g., the minimum score needed to predict proficiency).

Implications for Behavioral Difficulties

It is important to note that the three-tier MTSS, described above using examples to address achievement problems, is also appropriate for addressing behavior problems (National Association of State Directors of Special Education 2005). In fact, one of the essential attributes of an MTSS could be that it addresses both academic and behavioral difficulties with tiered interventions. Scholarship addressing the use of problem-solving models for behavioral difficulties has established an empirical foundation to build upon and many of the early MTSS models address behavior as well. For example, many point to Deno and Mirkin (1977) as the origin of both CBM and RTI, and the data-based decision-making process outlined in that seminal document was also applied to behavioral difficulties. This information is particularly important in the context of an increasing number of children in special education programs for children with emotional disturbance (U.S. Department of Education, 2003). Given the interplay of social, emotional, and behavioral adjustment with academic achievement, addressing problems in these areas is also critically important to enhance student success at school. The following provides a brief description of activities at each tier.

Tier 1

Universal screening is important to identify children at-risk or developing or displaying social, emotional, or behavior problems. Annual school-wide screening provides an opportunity for school-based professionals to better understand the student population and identify both individual students as well as systems-level areas of need. Some schools may administer brief student surveys or rating scales addressing social, emotional, or behavior problems to all students. In addition, students' cumulative records, teachers' gradebooks, or behavioral referral databases may be available in some schools. Another screening technique that may be used is to have all teachers identify children in their classrooms they are concerned about regarding specific social behaviors (e.g., peer relationship problems, inattention, poor classroom behavior, depression) (Demaray and Elliott 2001). In many schools, it is anticipated that such universal screening would reveal that 80–85 % of students would be in the healthy range (Walker and Shinn 2002). For those students identified at-risk or currently engaging in problem behaviors (e.g., affective problems, externalizing problems, social-relationship problems, risky behaviors), it is anticipated that core curriculum modifications would benefit many students. For instance, school-wide Positive Behavioral Interventions and Supports (PBIS; e.g., Sugai et al. 2000; Crone and Horner 2003) have been associated with promoting positive student behaviors.

Tier 2

For those students who continue to display problem behaviors, it would be essential to gather and carefully examine data for these individual students. The problem-solving process in tier 2 would involve both general and special education personnel, and is within the administrative and fiscal responsibility of general education. These data would be used to facilitate problem analysis and clearly define the problem(s). For instance, it would be important to identify the conditions under which the student is displaying problems and those where the student is not. These data can then be used for developing interventions for individual students or groups of students, as well as evaluating the effectiveness of those interventions, and may also be used to determine eligibility for special education services (Gresham and Project REACH 2005). Evidence-based interventions for small groups and individuals would be implemented in the classroom with appropriate evaluation to determine whether the problem behaviors were improving and if the interventions led to success on established behavioral objectives.

Tier 3

For those students who continue to display problem behaviors following tiers 1 and 2 prevention and intervention activities, it would be important to carefully review available data and discern whether additional information was necessary to understand why the problem behaviors persist. Assessment would include a comprehensive multidisciplinary assessment of the child's educational needs. The process should focus on gathering information that will help to clearly define the problem and facilitate the development of individual intensive interventions. Through the intervention process, data should be gathered repeatedly and often to monitor student improvement. Depending upon the student's RTIs during this third tier the student may or may not require additional support services.

If the results of this comprehensive evaluation indicate that a student's instructional needs cannot be met exclusively in the general education program, then an Individualized Education Program (IEP) team meeting would be convened to determine appropriate supports and services in special education.

Summary

Previous efforts of scholars and practitioners across the country provide a robust foundation of knowledge and insights to build upon to enhance the success of students through identifying and addressing their needs. Recent federal mandates and special education regulations serve as a catalyst to consider consolidating approaches rather than operating in a field of connected yet distinctly different models. The core components identified above may establish the basis from which to unify MTSS efforts and assure the implementation integrity that could be the most significant threat to a national movement. Moreover, using core components would enhance the likelihood of success and continue to enhance the educational success of children.

References

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research, 67,* 461–502.

Algozzine, B., & Ysseldyke, J. (1982). Classification decisions in learning disabilities. *Educational and Psychological Research, 2,* 117–129.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning. *Journal of Psychoeducational Assessment, 23,* 362–380.

Bransford, J., & Stein, B. (1984). *The ideal problem solver: A guide for improving thinking, learning and creativity*. San Francisco: W. H. Freeman.

Burns, M. K., & Gibbons, K. (2012). *Response to intervention implementation in elementary andsecondary schools: Procedures to assure scientific-based practices* (2nd ed.). New York: Routledge.

Burns, M. K., & Symington, T. (2002). A meta-analysis of pre-referral intervention teams: Student and systemic outcomes. *Journal of School Psychology, 40,* 437–447.

Burns, M. K., & VanDerHeyden, A. M. (2006). Using response to intervention to assess learning disabilities: Introduction to the special series. *Assessment for Effective Intervention, 32,* 3–5.

Burns, M. K., & Ysseldyke, J. E. (2005). Comparison of existing responsiveness-to-intervention models to identify and answer implementation questions. *The California School Psychologist, 10,* 9–20.

Burns, M. K., Vanderwood, M., & Ruby, S. (2005). Evaluating the readiness of prereferral intervention teams for use in a problem-solving model: Review of three levels of research. *School Psychology Quarterly, 20,* 89–105.

Burns, M. K., Ganuza, Z., & London, R. (2009). Brief experimental analysis of written letter formation: A case demonstration. *Journal of Behavioral Education, 18,* 20–34.

Burns, M. K., Karich, A., Maki, K., Anderson, A., Pulles, S. M., Ittner, A., McComas, J. J., & Helman, L. (in press). Identifying classwide problems in reading with screening data. *Journal of Evidence Based Practices for Schools*.

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *Communiqué, 34*(3).

Crone, D. A., & Horner, R. H. (2003). *Building positive behavior support systems in schools: Functional behavioral assessment*. New York: Guilford.

Demaray, M. K., & Elliot, S. N. (2001). Perceived social support by children with characteristics of attention-deficit/hyperactivity disorder. *School Psychology Quarterly, 16,* 68–90.

Deno, S. L. (2002). Problem solving as best practices. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 37–56). Bethesda: National Association of School Psychologists.

Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligence testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice, 13,* 186–203.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199–208.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13,* 204–219.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18,* 157–171.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention reading in the primary grades. A practice guide*.

Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.

Gickling, E. E., & Havertape, S. (1981). *Curriculum-Based Assessment (CBA)*. Minneapolis: School Psychology Inservice Training Network.

Gresham, F. M., & project REACH. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children, 28,* 328–344.

Gresham, F. M., Reschly, D. J., Tilly, W. D. III, Fletcher, J., Burns, M., Crist, T., Prasse, D., Vanderwood, M., & Shinn, M. (2004). Comprehensive evaluation of learning disabilities: A response-to-intervention perspective. *Communiqué, 33*(4).

Hall, M., & Burns, M. K. (2014). Meta-analysis of small-group reading interventions. Manuscript submitted for publication.

Howell, K., & Nolet, V. (1999). *Curriculum-based evaluation: Teaching and decision making*. Pacific Grove: Brooks/Cole.

Ikeda, M. J., Tilly, W. D. III., Stumme, J., & Volmer, L. (1996). Agency-wide implementation of problem-solving consultation: Foundations, current implementation, and future directions. *School Psychology Quarterly, 11,* 228–243.

Individuals with Disabilities Education Improvement Act. (2004). Pub. L. 108–446.

Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 113–142). New York: Guilford Press.

Kavale, K. A., Kauffman, J. M., Bachmeier, R. J., & LeFever, G. B. (2008). Response-to-intervention: Separating the rhetoric of self-congratulation from the reality of specific learning disability identification. *Learning Disability Quarterly, 31,* 135–150.

Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (2014). An introduction to universal screening in educational settings. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 3–17). Washington, DC: American Psychological Association.

Kovaleski, J. R., Tucker, J. A., & Stevens, L. J. (1996). Bridging special and regular education: The Pennsylvania initiative. *Educational Leadership, 53,* 44–47.

Kovaleski, J. R., Shapiro, E. S., & VanDerHeyden, A. M. (2013). *The RTI approach to evaluating learning disabilities*. New York: Guilford.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice, 18,* 187–200.

McDermott, K. A. (2011). *High-stakes reform: The politics of educational accountability*. Washington, DC: Georgetown University Press.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance standards on state assessments in reading. *School Psychology Review, 33,* 193–203.

National Association of State Directors of Special Education. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria: author.

National Reading Panel. (2000). *Report of the national reading panel: Teaching children to read*. Washington, DC: U.S. Department of Health and Human Services.

Petursdottir, A. L. (2006). *Brief experimental analysis of early reading interventions*. Doctoral Dissertation.

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families*. Washington, DC: US Department of Education.

Shapiro, E. S. (2000). School psychology from an instructional perspective: Solving big, not little problems. *School Psychology Review, 29,* 560–572.

Shapiro, E. S. (2010). *New thinking in response to intervention: A comparison of computer-adaptive tests and curriculum-based measurement within RTI.* Wisconsin Rapids: Renaissance Learning.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30,* 407–419.

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wise discipline interventions. *Journal of Emotional and Behavioral Disorders, 8,* 94–101.

Tilly III, W. D. (2002). Best practices in school psychology as a problem-solving enterprise. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 21–36). Bethesda: National Association of School Psychologists.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30,* 15–29.

VanDerHeyden, A. M., & Tilly, W. (2010). *Keeping RtI on track: How to identify, repair, and prevent mistakes that derail implementation*. Horsham: LRP Publishing.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32,* 204–227.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S., Chen, R., Pratt, A., & Denkla, M. B. (1996). Cognitive profiles of difficulty-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experimental deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

Walker, H. M., & Shinn, M. R. (2002). Structuring school-based interventions to achieve integrated primary, secondary, and tertiary prevention goals for safe and effective schools. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 1–25). Washington, DC: National Association of School Psychologists.

Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., Rosenfield, S., & Telzrow, C. (2006). *School psychology: A blueprint for training in practice III*. Bethesda: National Association of School Psychologists.

Index

S. R. Jimerson et al. (eds.), *Handbook of Response to Intervention,* DOI 10.1007/978-1-4899-7568-3