

ADVANCES IN ACCOUNTING BEHAVIORAL RESEARCH

Series Editor: Vicky Arnold

Recent Volumes:

Volumes 1–4: Edited by James E. Hunton

Volumes 5–10: Edited by Vicky Arnold

ADVANCES IN ACCOUNTING BEHAVIORAL RESEARCH VOLUME 11

ADVANCES IN ACCOUNTING BEHAVIORAL RESEARCH

EDITED BY

VICKY ARNOLD

*Ernst & Young Professor of Accounting,
Kenneth G. Dixon School of Accounting,
University of Central Florida, USA*

Professorial Fellow

*Department of Accounting and Business Information Systems,
The University of Melbourne, Australia*

ASSOCIATE EDITORS

B. DOUGLAS CLINTON

Northern Illinois University, USA

ANNE LILLIS

University of Melbourne, Australia

ROBIN ROBERTS

University of Central Florida, USA

CHRIS WOLFE

Texas A&M University, USA

SALLY WRIGHT

University of Massachusetts, Boston, USA



JAI

United Kingdom – North America – Japan
India – Malaysia – China

JAI Press is an imprint of Emerald Group Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2008

Copyright © 2008 Emerald Group Publishing Limited

Reprints and permission service

Contact: booksandseries@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. No responsibility is accepted for the accuracy of information contained in the text, illustrations or advertisements. The opinions expressed in these chapters are not necessarily those of the Editor or the publisher.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-84663-960-9

ISSN: 1475-1488 (Series)



Awarded in recognition of
Emerald's production
department's adherence to
quality systems and processes
when preparing scholarly
journals for print



INVESTOR IN PEOPLE

LIST OF CONTRIBUTORS

<i>Mohammad J. Abdolmohammadi</i>	Department of Accountancy, Bentley College, USA
<i>Asokan Anandarajan</i>	School of Management, New Jersey Institute of Technology, USA
<i>Réjean Belzile</i>	Department of Accounting, University of Quebec at Montreal, Canada
<i>Anthony P. Curatola</i>	Department of Accounting, Drexel University, USA
<i>Carlin Dowling</i>	Department of Accounting and Business Information Systems, The University of Melbourne, Australia
<i>Tim V. Eaton</i>	Department of Accountancy, Miami University, USA
<i>Dennis Eggett</i>	Department of Statistics, Brigham Young University, USA
<i>Amy M. Hageman</i>	Kenneth G. Dixon School of Accounting, University of Central Florida, USA
<i>James E. Hunton</i>	Department of Accountancy, Bentley College, USA
<i>Kip R. Krumwiede</i>	Department of Accountancy, Boise State University, USA
<i>Daniel W. Law</i>	School of Business Administration, Gonzaga University, USA
<i>Robyn Moroney</i>	Department of Accounting and Finance, Monash University, Australia
<i>Christine Nolder</i>	Department of Accountancy, Bentley College, USA

<i>Scott L. Summers</i>	School of Accountancy, Brigham Young University, USA
<i>Monte R. Swain</i>	School of Accountancy, Brigham Young University, USA
<i>John T. Sweeney</i>	Department of Accounting, Information Systems and Business Law, Washington State University, USA
<i>Chantal Viger</i>	Department of Accounting, University of Quebec at Montreal, Canada

REVIEWER ACKNOWLEDGMENTS

The Editor and Associate Editors at AABR would like to thank the many excellent reviewers who have volunteered their time and expertise to make this an outstanding publication. Publishing quality papers in a timely manner would not be possible without their efforts.

Mohammed Abdolmohammadi
Bentley College, USA

Craig Emby
Simon Fraser University, Canada

Elizabeth Dreike Almer
Portland State University, USA

Michael Favere-Marchesi
Simon Fraser University, Canada

John Anderson
San Diego State University, USA

Leonor Ferreira
Universidade Nova de Lisboa,
Portugal

Jean C. Bedard
Bentley College, USA

Timothy J. Fogarty
Case Western Reserve University,
USA

James Bierstaker
Villanova University, USA

Amy Hageman
University of Central Florida, USA

Richard Brody
University of New Mexico, USA

Karen L. Hooks
Florida Atlantic University, USA

Charles Cho
Concordia University, Canada

Susan Hughes
Butler University, USA

Carlin Dowling
The University of Melbourne,
Australia

James E. Hunton
Bentley College, USA

Kirsten Ely
Sonoma State University,
USA

Kevin Jackson
University of Illinois, USA

Karla Johnstone
University of Wisconsin Madison,
USA

Jennifer Kahle
University of South Florida, USA

Les Kren
University of Wisconsin Milwaukee,
USA

Jordan Lowe
Arizona State University, USA

Ted Mock
University of Southern California,
USA

Karen Nunez
North Carolina State University,
USA

William Pasewark
Texas Tech University, USA

Sean Robb
University of Central Florida, USA

Constantinos J. Stefanou
Alexander TEI of Thessaloniki,
Greece

Steve Sutton
University of Central Florida, USA

Bill Tayler
Emory University, USA

Mark Taylor
Creighton University, USA

Jay Thibodeau
Bentley College, USA

Greg Trompeter
Boston College, USA

Sally Widener
Rice University, USA

Brett Wilkinson
Baylor University, USA

Arnold Wright
Boston College, USA

EDITORIAL POLICY AND SUBMISSION GUIDELINES

Advances in Accounting Behavioral Research (AABR) publishes articles encompassing all areas of accounting that incorporate theory from and contribute new knowledge and understanding to the fields of applied psychology, sociology, management science, and economics. The journal is primarily devoted to original empirical investigations; however, literature review papers, theoretical analyses, and methodological contributions are welcome. AABR is receptive to replication studies, provided they investigate important issues and are concisely written, and is receptive to methodological examinations that can potentially inform future behavioral research. The journal especially welcomes manuscripts that integrate accounting issues with organizational behavior, human judgment/decision making, and cognitive psychology.

Manuscripts will be blind-reviewed by two reviewers and an associate editor. The recommendations of the reviewers and associate editor will be used to determine whether to accept the paper as is, accept the paper with minor revisions, reject the paper or to invite the authors to revise and resubmit the paper.

MANUSCRIPT SUBMISSION

Manuscripts should be forwarded to the editor, Vicky Arnold, at VArnold@bus.ucf.edu via e-mail. All text, tables, and figures should be incorporated into a word document prior to submission. The manuscript should also include a title page containing the name and address of all authors and a concise abstract. Also, include a separate word document with any experimental materials or survey instruments. If you are unable to submit electronically, please forward the manuscript along with the experimental materials to the following address:

Vicky Arnold, Editor

Advances in Accounting Behavioral Research

Kenneth G. Dixon School of Accounting

University of Central Florida

P. O. Box 161400

Orlando, FL 32816-1400, USA

References should follow the APA (American Psychological Association) standard. References should be indicated by giving (in parentheses) the author's name followed by the date of the journal or book; or with the date in parentheses, as in 'suggested by Canada (2005).'

In the text, use the form Hageman et al. (2006) where there are more than two authors, but list all authors in the references. Quotations of more than one line of text from cited works should be indented and citation should include the page number of the quotation; e.g. (Phillips, 2001, p. 56).

Citations for all articles referenced in the text of the manuscript should be shown in alphabetical order in the Reference list at the end of the manuscript. Only articles referenced in the text should be included in the Reference list. Format for references is as follows:

For Journals

Dunn, C. L., & Gerard, G. J. (2001). Auditor efficiency and effectiveness with diagrammatic and linguistic conceptual model representations. *International Journal of Accounting Information Systems*, 2(3), 1–40.

For Books

Ashton, R. H., & Ashton, A. H. (1995). *Judgment and decision-making research in accounting and auditing*. New York, NY: Cambridge University Press.

For a Thesis

Smedley, G. A. (2001). *The effects of optimization on cognitive skill acquisition from intelligent decision aids*. Unpublished doctoral dissertation, University.

For a Working Paper

Thorne, L., Massey, D. W., & Magnan, M. (2000). *Insights into selection-socialization in the audit profession: An examination of the moral reasoning of public accountants in the United States and Canada*. Working paper. York University, North York, Ontario.

For Papers from Conference Proceedings, Chapters from Book, etc.

Messier, W. F. (1995). Research in and development of audit decision aids. In: R. H. Ashton & A. H. Ashton (Eds), *Judgment and Decision Making in Accounting and Auditing* (pp. 207–230). New York: Cambridge University Press.

A REVIEW OF THE STRENGTHS AND WEAKNESSES OF ARCHIVAL, BEHAVIORAL, AND QUALITATIVE RESEARCH METHODS: RECOGNIZING THE POTENTIAL BENEFITS OF TRIANGULATION

Amy M. Hageman

ABSTRACT

This chapter discusses the benefits, limitations, and challenges in developing research projects that integrate a combination of archival, behavioral, and qualitative research methods. By demonstrating the inherent strengths and weaknesses of using a single method in isolation, this chapter aims to broaden our understanding of why and how research that examines various issues from the different perspectives is richer than employing any single method and enhances our understanding of a given accounting phenomenon. This chapter also discusses how investigating an issue through multiple research methods can help researchers improve the generalizability of findings and present a panoramic view of a particular phenomenon.

Advances in Accounting Behavioral Research, Volume 11, 1–30
Copyright © 2008 by Emerald Group Publishing Limited
All rights of reproduction in any form reserved
ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11001-8

INTRODUCTION

Accounting is an applied discipline, within which researchers rely on theories from a wide array of root disciplines to investigate research questions. Perhaps because of the variety of root disciplines that have been influential in helping to motivate accounting research (particularly economics, psychology, and sociology), accounting researchers employ a variety of research methods to conduct empirical studies of accounting phenomena. Some of the most common research methods can be broadly classified as *archival* (secondary sources of, mostly, numeric data at the organizational unit of analysis), *behavioral* (primary sources of, mostly, numeric data at individual, group, and organizational units of analysis), and *qualitative* (secondary and primary sources of, mostly, non-numeric data at individual, group, and organizational units of analysis).

Currently, researchers using archival research methods vastly outnumber those who use behavioral or qualitative techniques. A survey by [Koonce and Mercer \(2005\)](#) of five of the top accounting journals showed that from 1993 to 2004, over 94% of financial accounting studies used archival research methods, compared to less than 6% that employed behavioral methods. Similarly, a more comprehensive survey of 14 accounting journals from 1981 to 2004 by [Merchant and Van der Stede \(2006\)](#) found that very little accounting research used field research and other qualitative techniques. Part of the reason for the dominance of archival research methods concerns the teaching philosophies of many accounting doctoral programs – whereas virtually all U.S. accounting doctoral students take at least one course emphasizing archival research methods, fewer than half are typically exposed to behavioral methods ([Koonce & Mercer, 2005](#)) and only a small number are trained in qualitative research techniques ([Merchant & Van der Stede, 2006](#)).

Equally troubling is that doctoral students' training often emphasizes the *research method* rather than the *research question*. As stated by [Koonce and Mercer \(2005, p. 177\)](#), “Accounting doctoral students typically choose one of two, largely non-overlapping, fields of specialization ... this choice is often determined by the type of method the student expects to use in his or her coursework [archival or behavioral].” Appropriate research methods should certainly match a study's research questions. However, given that most future accounting researchers are trained to examine a problem within the confines of a particular research method, most researchers rely on a singular method in examining their research questions, and limit their views of their particular paradigm based on the research method used

(see Burrell & Morgan, 1979). This limitation is perpetuated by the inherent difficulties in developing and maintaining skill sets in multiple methods.

The purpose of this chapter is to discuss the benefits, limitations, and challenges in developing a well-designed study using any single research method in isolation. In particular, raising accounting researchers' awareness of the limitations of their "dominant" research method is important, as it can help to foster deeper understanding of the potential contributions of other techniques. While researchers within the management accounting discipline have long recognized the importance of using integrative research methods (e.g., the call for cross-sectional field study research in Lillis & Mundy, 2005), this chapter strives to show how *triangulation*¹ of research methods (using two or more research methods to study a given topic) can produce a richer, more complete understanding of all accounting phenomena. Currently, the academic accounting community emphasizes the use of archival research methods at the expense of behavioral and qualitative research. By discussing the shortcomings in each of these types of research methods, this chapter aims to illustrate how multiple papers on the same research topic should use differing methods in order to overcome the weaknesses in collectively relying upon a single approach to research.

This chapter is especially aimed at less-experienced academics (i.e., doctoral students and new faculty members) as such individuals can benefit from understanding how each research method presents its own set of challenges, and why investigations of accounting phenomena through multiple methods helps present a more comprehensive view of a specific phenomenon. The discussion of alternative research techniques within this chapter can also be informative to more experienced researchers who wish to investigate research questions using methods that are outside of the focus of their initial training.

Many previous authors have certainly noted the strengths and weaknesses of archival, behavioral, and qualitative research methods in the social sciences in general (e.g., Kerlinger & Lee, 2000; Shadish, Cook, & Campbell, 2002; Berg, 2003; Greene, 2003) and in accounting research in particular (e.g., Kothari, 2001; Libby, Bloomfield, & Nelson, 2002; Merchant & Van der Stede, 2006). While all of these articles and textbooks constitute important examinations of specific research methods, none provides a combined overview of *all* of these different methods in the context of accounting research. This chapter therefore aims to synthesize information on the benefits and limitations of research methods as used in accounting research within the framework of a single paper, which overviews the

research methods field and provides new scholars with a single resource for beginning their investigations of how to conduct research.

The remainder of this chapter is organized as follows: The next three sections detail the challenges involved in conducting a well-designed study using archival, behavioral, and qualitative research methods. The subsequent section provides a brief example of how examining a research topic with multiple methods can help produce a more complete understanding of the underlying accounting phenomena. The final section concludes with a discussion of the importance of triangulation in accounting research.

ARCHIVAL RESEARCH METHODS

Strictly speaking, the use of an “archival” research method entails the use of secondary data sources, in which researchers analyze data contained in an archived record. However, as commonly used in the accounting discipline, “archival” denotes the use of mostly *numeric* data; in practice, many studies employing this method use large-scale, secondary numerical data in the positivist tradition (Burrell & Morgan, 1979).² This type of research method was not common in academic accounting research until the paradigm-shifting work of Ball and Brown (1968) and Beaver (1968), and today represents the dominant research method of first choice when conducting accounting research (see Watts & Zimmerman, 1986; Kothari, 2001; Koonce & Mercer, 2005).

The archival research method is particularly useful in its ability to examine trends in large-scale data. Thus, external validity is particularly high in studies using archival research methods, as such studies use data pertaining to *naturally occurring* events. The archival research method is particularly appropriate for examining *macro-level patterns* (broad economic or societal trends), such as general economic trends over time, but is also commonly used for examining micro-level behavior in the *aggregate*. Many archival studies therefore use *econometrics*, or “the application of mathematical statistics and the tools of statistical inference to the empirical measurement of relationships postulated by economic theory” (Greene, 2003, p. 1). This analysis is useful to accounting researchers who examine large-scale trends of naturally occurring events, such as the stock market’s reaction to a new accounting standard.

Despite the usefulness of archival research methods, challenges remain. These relate to the general problems in using secondary numerical data sources to draw causal inferences, as well as to the more mundane nuances associated with conducting a quality archival study.

Difficulty in Establishing Causal Inferences

One of the primary challenges associated with archival research is that researchers *cannot easily test causal relationships*. Relying on secondary data often compromises an archival study's internal validity. The researcher can analyze data trends, but it is difficult to establish that a particular factor causes another. In most archival studies, researchers examine the relationships between certain variables and attempt to control for other alternative explanations that may have affected the relationship. Thus, archival research designs have relatively low levels of internal validity, as it is difficult for the researcher to properly control for all other plausible explanations for an observed relationship between phenomena. This is particularly true for *non-experimental designs*, which lack randomization, control groups, pre-tests, and other factors; researchers instead measure and statistically control for alternative explanations (Shadish et al., 2002). Establishing causality with archival studies is especially challenging with *cross-sectional studies*, in that it is unclear whether the purported cause actually precedes the event; it is therefore critical that all other potential causal factors are well-measured and controlled, and that the model is very well specified (Shadish et al., 2002).

Since archival research cannot clearly indicate causality, *endogeneity* of independent variables is another common concern. Researchers may test the effect of an independent variable on a dependent variable, but ignore the antecedents of the independent variable itself. An analyst's affiliation is one such endogenous construct, as companies may choose analysts that are naturally optimistic (Kothari, 2001).

Archival researchers attempt to address the difficulties of establishing causal inferences in several ways. First, since establishing causality in archival research is difficult, the researcher must clearly *control for and explain other alternative explanations*. This manner of helping to mitigate threats to internal validity often requires the researcher to become immersed in statistical methods and techniques, since controlling for other explanations can be challenging.

Researchers conducting archival research may encounter difficulties when the model violates assumptions of multivariate analysis, such as normality, homoskedasticity, and linearity (Hair, Anderson, Tatham, & Black, 1998). Researchers who are aware of these common problems may try to transform their data before the analysis, such as *scaling variables* by the size of assets in order to help eliminate heteroskedasticity or by taking the natural log of a variable (such as audit fees in Simunic, 1980). However, all of these

statistical techniques represent a high barrier to entry for those not well trained in these methods, as those not well trained in statistical techniques are unlikely to succeed in conducting the extensive statistical testing and analysis required to assure the reliability of the results.

A second way of alleviating the problem in archival research of establishing causal inferences is the use of *natural experiments*, in which the researcher examines the relationship between a naturally occurring event and a comparison event (Shadish et al., 2002). Natural experiments help address causality concerns by providing built-in temporal precedence of the proposed cause in relation to its effect. Many archival studies of earnings management use natural experiments in testing their hypotheses, such as studying earnings management among firms undergoing or not undergoing import price relief investigation (Jones, 1991). Moreover, many econometric studies have shifted toward the use of natural experiments as a tool for establishing valid inferences; the use of this technique is particularly attractive in the econometric domain since researchers cannot easily manipulate variables of interest (Shadish et al., 2002).

A third compensating technique used by archival researchers is the *matching of observations*. In this case, the researcher is not able to achieve random selection and random assignment, but can match observations on certain characteristics. This method is not foolproof, as it is very difficult to know which dimensions should be matched, or even that an exact match has actually been made, but it helps to alleviate some of the problems with pure non-experimental designs by assuring that the control and treatment groups have equivalent matched dimensions (Kerlinger & Lee, 2000). Some studies of earnings management have used matching techniques. For example, Holthausen, Larcker, and Sloan (1995) studied the relationship between management bonus plans and manipulation of earnings, and matched firm managers' performance on bonus plans characteristics; Klein (2002) used a matched portfolio matched by abnormal accruals in examining the relationship between earnings management and corporate governance structures. The differences between matched pairs help to establish some of the relevant antecedents of earnings management.

Difficulty of Secondary Data Reliance

Another major difficulty in conducting quality archival research arises from this method's reliance upon secondary data sources. This reliance brings about a host of additional problems that archival researchers must address.

Measurement error of the variables is abundant with the use of secondary data sources and is one of the chief difficulties in using archival research methods to test the relationship among variables in a model. Specifically, threats to construct validity are particularly prevalent in archival research, as sampled variables rarely are exact proxies for the latent theorized constructs. Indeed, Greene (2003, p. 8) writes that, "... the difficulty of obtaining reasonable measures of profits, interest rates, capital stocks, or, worse yet, flows of services from capital stocks is a recurrent theme in the empirical literature."

Measurement error due to poor construct validity arises from two related reasons – the data may be very poorly measured, or data that is needed to operationalize a theoretical construct may simply be unavailable in a secondary dataset. Poorly measured data is a common concern in archival studies. For instance, earnings management studies may use the Jones (1991) model of discretionary accruals to approximate the accruals that management may manipulate, but since this construct is unobservable to those outside of the company, the approximation of "discretionary accruals" contains a great deal of error. A related challenge is that some variables may be inherently immeasurable in numeric form, hence, not available in an existing dataset. Archival studies that attempt to measure constructs such as investors' "expectations" about future stock market prices (see Greene, 2003) suffer from extreme measurement error, since no secondary dataset truly captures "expectations" and the construct of expectations itself is difficult to describe numerically. Thus, nearly all archival studies in accounting doubtlessly suffer from measurement error due to threats to construct validity, and empirically observing the proposed theoretical relationship may be difficult. Researchers acknowledge these threats to construct validity by carefully selecting their operationalized variable from an existing data set and disclosing limitations.

The reliance on secondary data is particularly problematic if there are *errors in the datasets themselves*. For instance, AuditAnalytics is a popular dataset containing information on auditing information by company such as corporate auditor, auditor fees, and Sarbanes–Oxley (SOX) disclosures and compliance. However, investigation of this dataset has shown that there are numerous inconsistencies between the information reported in AuditAnalytics and in the actual 10-K filings (Canada, Kuhn, & Sutton, 2008). This means that prior studies using information from AuditAnalytics may have relied on incorrect data. Alternatively, information may not match across datasets. For instance, unlike Compustat, I/B/E/S adjusts earnings for one-time events and special items (Kothari, 2001).

Whether researchers are cognizant of these discrepancies or make the necessary transformations or adjustments is not always clear.

Another problem with relying on secondary datasets is that much of this research, particularly capital markets events studies, uses secondary data from a *long time period* in order to enhance the generalizability of the findings. However, with longer time horizons, controlling for all confounding factors is difficult. Another problem particularly prevalent in capital markets research is survivorship bias; observations not available for all periods are excluded from the analysis, which in capital markets research biases the results toward older, more successful firms. One way to solve the problem of data availability is to use data available at greater intervals, such as quarterly earnings announcements; quarterly data helps increase the power of statistical tests, alleviate survivorship bias, and can potentially expose seasonality (Kothari, 2001).

Other problems arise from the use of *large-scale data*, particularly in capital markets studies. Variables in these studies may be serially dependent or biased by correlated omitted variables. These concerns may be mitigated by the use of *first differences*, in which data is lagged over a period; some researchers may choose to lag the data for additional periods. Controlling for cross-correlation in the data may also address these concerns (Kothari, 2001). With the use of large-scale data, archival researchers often practice data truncation, whereby outliers are removed from the analysis. This practice biases the results toward average values and away from the extremes.

Archival researchers must also contend with a number of *practical data limitations*. First, researchers that utilize publicly available secondary data, such as information from databases such as CRSP or Compustat, must compete with a large pool of other researchers also using the same data source. It may be harder to develop an original testable hypothesis that uses these widely held databases. Second, researchers that use data from a propriety source (such as Holthausen et al., 1995 and its use of a propriety compensation database) must incur large monetary or temporal expenses. This ensures that researchers may need a variety of resources to conduct high-quality archival research. Third, researchers are limited by the availability of data; some data is only available in limited periods. For instance, Omer and Shelley (2004) were unable to test their tax competition hypothesis before 1978 because a great deal of state-related data was not available before this time.

Finally, archival studies are limited by what has *actually occurred*. In addition to the fact that this enhanced external validity comes at the cost

of reduced internal validity, archival studies are not as useful in testing the consequences of proposed policy action or regulatory change.

Overall

Overall, archival research methods are very useful in helping researchers examine macro-level patterns in naturally occurring events, but suffer from poor internal and construct validity, along with a host of other problems associated with using large-scale secondary data sources. In particular, the use of questionable proxies in pure archival research often limits researchers' attempts at understanding a phenomenon. Triangulation of archival methods with behavioral and qualitative research methods may help alleviate some of these concerns.

BEHAVIORAL RESEARCH METHODS

Behavioral research methods are particularly useful in understanding human behavior. These methods can focus on individuals, groups, or organizations as units of analysis. "Behavioral" research is a broad umbrella encompassing a variety of research methods, including laboratory experiments (where participants engage in contrived tasks under controlled, yet artificially created, experimental conditions), surveys (where respondents reflect their beliefs, attitudes, cognitions, and motivations under general conditions), experimental economics (where buyers and sellers interact in an artificial market setting), and field experiments (where participants work in their natural environments under various conditions of interest). However, as used in the academic accounting paradigm, the term "behavioral" typically denotes a laboratory experiment. Despite the use of behavioral research methods across all functional areas of accounting, however, the use of this research method is not as common as the archival method (see [Koonce & Mercer, 2005](#)).

The single greatest advantage of behavioral research that uses laboratory experiments is the ability to *establish causality* in relationships between and among phenomena. As stated by [Shadish et al. \(2002, p. 18\)](#), "The strength of experimentation is its ability to illuminate causal inference." This is because in an experiment, the researcher can isolate and control for other potentially confounding factors, which provides evidence against alternative explanations. To establish causality, researchers need to demonstrate

temporal precedence (that a cause occurred before an effect), co-variation (that the cause and effect are correlated), and that alternative explanations can be ruled out (Shadish et al., 2002). Thus, a great benefit of experiments is rooted in the concept of control. This offers the experimental method a high degree of internal validity in that tightly-designed and well-executed experiments can illustrate causal relationships. While experiments in which the independent variables are manipulated and participants are randomized among experimental conditions have the highest degree of internal validity, quasi-experiments and experiments that measure the independent variables can still investigate causality to some extent.

In general, behavioral methods are also useful for helping to understand *individual differences*. Some archival studies seek to examine questions of individual behavior (such as how analysts incorporate information, e.g., O'Brien, 1988), but behavioral methods are particularly effective in shedding light on individual-level phenomenon. This is especially the case in studies of factors for which the use of archival methods is impossible, such as understanding knowledge structures or other cognitive elements (e.g., Nelson, Libby, & Bonner, 1995). In that sense, behavioral methods are often superior to archival methods in their ability to use *primary source data*.

In summary, Kerlinger and Lee (2000) characterize laboratory experiments (the most common behavioral method used in accounting research) as having the primary strengths of control, randomization, precision, and manipulation and measurement. Because of these benefits, a true experiment is considered a scientific ideal, particularly when testing and building theory. However, laboratory studies, experimental economics, surveys, and field studies all have a host of limitations.

Laboratory Experiments

An informal survey of the journal *Behavioral Research in Accounting* indicates that laboratory experiments are the most common research method used by behavioral accounting researchers. While there are benefits from using laboratory experiments, they also have their own set of specific challenges.

One of the greatest strengths of the laboratory experiment is its high degree of control, such that researchers can learn about human cognitions and behavior by systematically varying certain factors and controlling for alternative explanations. Ironically, this very strength is also one of the primary challenges faced by behavioral experiments. In order to carry out

studies with a high degree of internal validity, researchers must conduct studies in *contrived, artificial environments* that remove participants from their natural settings. This process results in a loss of external validity and can hamper the generalizability of the study. It is difficult for a tightly designed experiment to also tap into the richness of the natural environment. For example, [Ashton \(1990\)](#) is an example of a study with a high degree of internal validity (all of the independent variables are tightly manipulated), but in which the generalizability is compromised (for instance, “incentives” in this study were operationalized as a tournament-type scheme that would not apply to practicing auditors). In summary, high internal validity and control come at a cost, the lack of generalizability and external validity; therefore, researchers must exercise caution when generalizing the results of laboratory studies to non-laboratory settings. Researchers must balance internal and external validity concerns when designing experiments. However, it must be reinforced that the main purposes of experiments are to test and build theory, and that external validity can be achieved to some degree by extrapolating experimental results through theory, not through tasks or settings.

A related challenge is that laboratory manipulations often produce relatively *weak effects* ([Kerlinger & Lee, 2000](#)). This is related to the fact that the laboratory is a contrived environment, where reproducing the precise effects that may be seen in the natural setting is quite difficult. [Kennedy \(1993\)](#) is an example of a weak manipulation; the study tested order effects, but auditor participants did not exhibit order bias in the experiment. In order to determine whether the manipulation has the intended effect, experiments generally include manipulation checks to ensure that participants view the experimental treatment in the same light as the researcher. Behavioral researchers typically conduct extensive pre-tests to ensure that the manipulations will work, and make changes to the experimental design if effects are not detected. In most instances, participants that do not pass the manipulation check are excluded from further analysis, although some studies only examine whether participants in experimental conditions passed the manipulation check in the aggregate. Researchers must carefully design experiments to avoid the limitations of weak effects.

Beyond the limitation of weak effects, laboratory experiments *cannot estimate the magnitudes of effects*. In an experiment in which the researcher manipulates one or more independent variables, the results can provide information on whether an effect occurred, but it is difficult to quantify the absolute magnitude of such an effect. Thus, it is safe and appropriate for experimental researchers to interpret their findings as ordinal, not interval, effects.

Laboratory experiments often require participants to perform a *task* during the experiment, which means that the researcher is responsible for developing an accurate, realistic task that can be used to test the study's hypotheses. Developing a task that is both appropriate for participants and can be used to adequately test the study's hypothesis is challenging, particularly when the study is examining a sensitive issue. For example, [Hunton, Libby, and Mazza \(2006\)](#) required participants to determine which available-for-sale security to sell as a proxy for earnings management for the study's investigation of whether more comprehensive disclosures help to mitigate earnings management behavior. While the ability to generalize about earnings management behavior within the confines of this particular task is somewhat limited, through theory, their study does indicate that when given sufficient discretion in accounting standards and personal incentives, managers will likely use such discretion to achieve personal gains. On a related note, determining how to *operationalize the study's constructs* of interest may be difficult. Behavioral studies often examine individual behavior in terms of psychological constructs that may be difficult to define or properly measure, such as "procedural knowledge" (e.g., [Bonner & Walker, 1994](#)).

Furthermore, proper laboratory and other behavioral studies require adequate *participants* who are an appropriate match for the experimental tasks. Accounting studies in all functional areas tend to use practitioners when appropriate, such as the use of auditors for much of the judgment and decision-making literature stream ([Gramling, Johnstone, & Mayhew, 2001](#)). This means that researchers must recruit practitioners, which can require significant amounts of time and money, and may be difficult if CPA firms refuse to cooperate due to liability or other concerns. Unlike the psychology discipline, accounting behavioral studies have access to a limited pool of potential participants (see [Gibbins, 1992](#)), so it is very difficult to replicate an experiment if something goes amiss. Due to the difficulties in obtaining accounting practitioners as participants, many accounting researchers conduct experiments with undergraduate or graduate students. While the use of student subjects may be appropriate if the experimental task is properly constructed ([Elliott, Hodge, Kennedy, & Pronk, 2007](#)), many studies relying upon student participants do not satisfy this criterion (see discussion in [O'Neil & Samelson, 2001](#)).

Experimenters must also ensure that participants are properly *motivated* and are engaged in the task. Subjects that are not properly motivated by the degree of experimental realism or via attention to the task will not produce usable data. Some researchers use incentives in order to increase attention to

the task at hand; however, using incentives to motivate participants can introduce its own series of problems (Bonner & Sprinkle, 2002). Overall, there is extensive pressure in each experiment to make sure that all procedures are properly implemented.

Another challenge for the researcher is that participating in an experiment exists as part of a *social situation* (Shadish et al., 2002). Within a laboratory experiment, participants may respond differently due to social cues from the researcher. Likewise, many experimental findings are dependent upon participants' perceptions of the task or of measurement items. This means that researchers must rely upon individual-level data that is often perceptual and may not adequately measure the construct of interest (Kerlinger & Lee, 2000).

These inherent limitations in experiments lead to the constraint that most experiments also require *extensive work prior to the collection of data*, through theoretical development, experimental design, task development, pre-tests, human subjects' approval, and so forth. Since a tight experimental design is critical to achieving experimental control, researchers must ensure that their task, measurements, and procedures are perfectly developed. While all research methods require an extensive, pre-data collection time commitment, making changes to the research design after data collection has begun is extremely difficult when using behavioral research methods. This restriction can add to considerable development time; for instance, the software *Insolve*, which is used in many experiments that study insolvency practitioners' judgments (e.g., Arnold, Collier, Leech, & Sutton, 2000), took nearly seven years to develop (Leech, Collier, & Clark, 1998). Furthermore, experimenters also need to control for the fact that human information processing is subject to bias (Tversky & Kahneman, 1974). As an example, individuals are influenced by the order of presentation (Hogarth & Einhorn, 1992), so experimenters must take care to randomize the presentation of experimental materials. Taken together, the time to develop an experiment and the factors that must be considered are quite challenging.

Finally, experiments are not immune from problems with *measurement error*, including threats to construct and internal validity. These threats are especially prevalent for the experimental variables that are challenging to manipulate in the laboratory and are instead measured. Many studies that examine "experience," for example, tend to measure participants' years or level of experience (e.g., Tubbs, 1992). While this measurement can be surmounted (for instance, Hampton, 2005), independent variables that are measured provide weaker results than those that are manipulated, and

may not necessarily be appropriate proxies for the higher-level constructs of interest.

Experimental Economics

Experimental economics may be thought of as a special type of laboratory experiment. In this research method, experimenters create an artificial market in order to study the behavior of buyers and sellers. Some experimental economics studies utilize multiple periods to study how behavior may change over time (e.g., Sprinkle, 2000). This specificity of focus is one of the primary strengths – researchers are able to reproduce a marketplace to determine how purchasers and vendors behave. Many of these studies rely on student participants, as students are appropriate proxies for vendor and consumer behavior.

Like the laboratory experiment, experimental economics benefits from very tight experimental design and high internal validity. This high internal validity comes at a price of *reduced realism and generalizability*. Indeed, some commentators have criticized experimental economics studies for developing such a low level of experimental realism as to be practically meaningless. An example is criticisms of studies of taxpayer behavior using this method and removing all social and moral contexts (O’Neil & Samelson, 2001).

Practically, experimental economics studies are *expensive* to run. Since most of these studies examine buyer and seller behavior within an artificial market, researchers must provide funds to mimic a marketplace; many studies guarantee a minimum payment to all participants, but the determination of how much each participant is awarded depends on his/her performance (e.g., Sprinkle, 2000). Given that this method is expensive to conduct, the researcher must ensure that the experiment is perfectly executed.

Surveys

The use of the survey method uses participants to respond to questionnaires; these questionnaires contain scales with items that measure certain constructs of interest. There is no manipulation involved; instead, the researcher measures all of the variables of interest. A review of *Behavioral Research in Accounting* indicates that the use of surveys has waned in most

recent behavioral accounting research, although their use continues to be more popular in the management accounting domain (as evidenced by an examination of management accounting journals such as *Journal of Management Accounting Research* or *Management Accounting Research*). Still, surveys do have some advantages over experiments, particularly in that they may be cheaper to conduct and may garner a higher degree of realism and external validity than found in an artificial laboratory environment. A well-administered survey may also generate a greater volume of information than in a single experiment (Kerlinger & Lee, 2000).

One of the greatest challenges associated with survey research, however, is that the data collected from this method is entirely *perceptual*. Thus, the value of survey data is only as valid as the reporting of participants' perceptions. Many people may not respond, tell the truth, or know the answer to the question, which further biases the results. Some techniques help to alleviate these concerns. For instance, following the techniques of survey design in Dillman (2006) can help to alleviate non-response bias, whereas techniques such as the randomized response technique (see Bailey, Hasselback, & Karcher, 2001) can help compensate for individuals' tendencies to answer in an inaccurate, socially desirable manner. Nevertheless, the information provided by survey respondents is often impossible for researchers to verify (Kerlinger & Lee, 2000); and, these self-reports may threaten the study's construct validity (Shadish et al., 2002).

Practically, even when attempting to control for non-response bias, most surveys still suffer from *low response rates*; response rates of around 10% are not uncommon when randomly surveying accounting practitioners nationwide (e.g., Bobek & Radtke, 2007). Furthermore, like an experiment, a survey is a social event, in which results may suffer from *social desirability bias* (participants respond in a manner that is socially desirable, but does not reflect their actual outcomes or attitudes).

Field Experiments

The final primary category of behavioral methods is the field experiment, in which researchers undertake a company investigation to study relationships among variables. Typically, researchers utilize a realistic setting and manipulate an independent variable to the degree allowed by the organization or other setting (Kerlinger & Lee, 2000). This type of research design is similar to a laboratory experiment, but is typically carried out in a more *realistic*, naturalistic environment. This type of method enables the

examination of *richer*, more complex problems than seen in laboratory experiments.

Despite these strengths, several challenges remain in conducting reliable field experiments. The first is that field experiments have a lower degree of internal validity than laboratory experiments, because “the independent variables are contaminated by uncontrolled environmental variables” (Kerlinger & Lee, 2000, p. 582). This means that the researchers should implement as many controls as possible over other potential variables to be assured of the reliability of the results. Furthermore, true *randomization* is very difficult in a field experiment. Many companies may tend to resist having their employees randomly assigned to conditions; therefore, in most field experiments, the independent variable of interest cannot be disentangled from other potential factors. Threats to internal validity due to field experiments’ lack of control introduce more random noise in the results than in laboratory experiments.

Overall

Overall, behavioral research methods attempt to collect primary data from individuals in order to gain knowledge about human behavior. Whereas all methods grant the researcher much more flexibility in investigating research problems than do archival methods, all types of behavioral research methods suffer from challenges. In short, because researchers collect data from individuals in behavioral studies, all data is in some sense self-assessed. While some methods have more control over this than others (i.e., laboratory experiments are better controlled than surveys), all research is dependent upon individual responses for data. Thus, behavioral and archival research methods can complement each other.

QUALITATIVE RESEARCH METHODS

Qualitative research methods rely mostly on non-numerical data; Kerlinger and Lee (2000, p. 588) formally define this method as, “... social and behavioral research based on unobtrusive field observations that can be analyzed without using numbers or statistics.” As compared to archival and behavioral studies, qualitative studies use an interpretive approach that is more focused on understanding *meaning* than on assessing causal

relationships. Thus, qualitative research is often used to investigate more complex “how” and “why” research questions.

Qualitative research is a broad category with research traditions in fields such as sociology, anthropology, or education (Kerlinger & Lee, 2000). As used in academic accounting, qualitative studies of data are typically either *field studies* (involving direct contact with real-world participants) or a *content analysis* (involving non-numerical analysis of primary or secondary communications). While this research method is less common in the North American academic accounting community and is not generally taught in doctoral education programs (Merchant & Van der Stede, 2006), its use is generally more accepted in European and Australian universities.

The greatest advantage of qualitative research methods is in the ability to analyze *naturalistic environments* (Merchant & Van der Stede, 2006). Such studies are able to examine more complex research questions, without sacrificing the *richness* or complexity of the natural environments. The process of carrying out qualitative research also has a high degree of flexibility; in most qualitative studies, the research questions and coding schemes can change during the research process (see Berg, 2003). Researchers can also select which samples to present at the end of their data collection period, rather than in the beginning (Kerlinger & Lee, 2000). While these characteristics of qualitative studies help strengthen the research, these same characteristics can also represent methodological challenges.

Qualitative Research in General

One challenge to qualitative research is in mitigating the potential challenge of *researcher bias*. Whereas bias can affect research design in all methods, qualitative research methods are particularly prone to this challenge due to the potentially subjective nature of data collection and analysis. However, qualitative researchers argue that if the qualitative researcher is a well-trained observer, then, “If done properly, the data collected from qualitative research can yield more information and less spurious variability than other research methods” (Kerlinger & Lee, 2000, p. 590). Thus, the training of the researcher to function as unobtrusively as possible and the disclosure by the researcher of all of his or her assumptions can help to mitigate this concern. As an example, many qualitative researchers will quote liberally from their interview notes or consulted text in order to assemble evidence in support of propositions (e.g., Greenwood, Suddaby, & Hinings, 2002).

Related to the potential of researcher bias is the concern that qualitative data requires much more *interpretation* on the part of the researcher. As an example, Anderson-Gough, Grey, and Robson (2005) used in-depth interviews of audit trainees to help understand the embedding of gender relationships in CPA firms. Their analysis required considerable interpretation in the developing of themes (e.g., temporal commitment) and the assignment of portions of interviews with informants to these themes. The challenge of interpretation can be addressed by using the technique of “analytic interpretation” that attempts to find a negative case within the set of hypothesized relationships (Merchant & Van der Stede, 2006), or to assume initially that there are no relationships among the data in order to evaluate emerging patterns with an open mind (Berg, 2003). Problems with interpretations of construct measurement can also be mitigated by using multiple coders of the study’s data.

Explaining and predicting causal relationships with the use of qualitative research is difficult. Whether this is perceived as a weakness of the method depends on the researcher’s assumptions. Many researchers, particularly those from a functionalist paradigm, view the difficulty in investigating causal relationships in qualitative studies as a serious deficiency, since the data examined in most qualitative studies is too complex to garner cause and effect relationships (Shadish et al., 2002). However, other researchers do not see qualitative research as striving to explain or predict relationships among phenomena, but see its goal as helping to explain and understand constructs, such as the nature of the “public interest” (Baker, 2005). Researchers wishing to explain causal relationships would probably be better advised to use an experimental method.

From a pragmatic standpoint, qualitative research entails numerous *ethical issues*; confidentiality and acting with high ethical conduct are particularly important when the researcher is personally involved with the study’s participants (Kerlinger & Lee, 2000). Conducting qualitative research is also extraordinarily *time intensive*. Some research may take years solely to gather data; for example, Broadbent and Laughlin (1997, 1998) conducted in-depth investigations of physicians’ and schools’ responses to changing accountability requirements, and collecting interview data initially took several years to complete. Furthermore, researchers often need organizational contacts in order to obtain access to certain types of information, such as historical records (e.g., Chua & Poullaos, 2002) or interviews with audit partners (e.g., Greenwood et al., 2002). This type of extensive organizational cooperation may not always be forthcoming.

Accounting research that uses qualitative methods generally employs either field studies or content analysis of secondary narrative documents. In addition to the general concerns outlined above, each of these specific methods contains its own unique set of challenges.

Field Studies

A field study is similar to a field experiment, but does not employ randomization, manipulation, treatment groups, control, or any other elements that are similar to laboratory experiments. Merchant and Van der Stede (2006, p. 118) specify that field studies entail the “in-depth study of real-world phenomena through direct contact with organizational participants.” Kerlinger and Lee (2000) treat field studies as involving non-experimental methods that seek to discover relationships and interactions present in real social structures. Both of these perspectives emphasize that field studies are *naturalistic* investigations in which the researcher acts as an observer and investigator. Field research is particularly helpful in understanding the *context* in which events occur. Within accounting, the use of this method has increased in the past few decades, but its use is still primarily confined to managerial accounting (Merchant & Van der Stede, 2006). This type of investigation can be particularly helpful in building theory.

In addition to their unique strengths, field studies also face several unique challenges. One is the process of *field selection* (sample selection). Researchers must carefully choose fields that accomplish the goals of the study, such as choosing a mix of organizations that have experienced both success and failure in the implementation of enterprise systems (e.g., Nicolaou, 2004). Since field studies often use analyses of a small number of organizations, proper field selection is critical.

The next challenge deals with *generalizability* and external validity. Qualitative field studies tend to use much smaller samples, but study the chosen samples in much greater depth in order to generalize the study’s findings to *theory*. This type of analysis permits the user to gain a richer understanding of a particular field’s context, but potentially sacrifices the ability to generalize the findings to other contexts. For example, Seal, Berry, and Cullen (2004) used a case study to attempt to theorize about the interorganizational role of the supply chain, but only used a single firm for analysis. This limits the ability to apply findings from the study to subsequent works.

Another specific limitation of field research is that within North America, many accounting academicians are not well trained in how to conduct quality field research. In particular, most novice field researchers do not tie in theory or prior accumulated knowledge in their analysis of results (Merchant & Van der Stede, 2006). This limitation can be overcome with specific training in field research techniques.

Finally, the richness in a field study may make it difficult for a researcher to consider all of the possible variables and factors that may affect an outcome. For example, a field study on interorganizational networks may provide rich detail that networks tend to function like enterprises, but cannot precisely specify all of the factors that may influence this result (e.g., Mouritsen & Thrane, 2005). The goal of most qualitative research is therefore to understand, rather than to predict and explain (Llewelyn, 2003).

Content Analysis

Another common type of qualitative method involves the content analysis of secondary, non-numerical communications. This type of method is unobtrusive, cost-effective, and particularly useful for studying long historical trends or general societal changes (Berg, 2003). For instance, Chua and Poullaos (2002) used historical documents to study the relationship between professional associations at the center and periphery of empire in the 19th century.

Like all other research methods, content analysis has its drawbacks. According to Berg (2003, p. 288), “The single serious weakness of content analysis may be in locating unobtrusive messages relevant to the particular research questions.” This is particularly a challenge when content analysis is used as the primary research method. The challenge of *locating unobtrusive messages* arises because the documents analyzed have already been recorded and do not contain data pertaining to the initial research question. Therefore, performing a content analysis on research questions such as the history of professional associations in Trinidad and Tobago (Annisette, 2000) requires the researcher to extensively search documents in order to develop themes related to the research question.

Secondly, content analysis is “virtually useless” in investigating causal relationships (Berg, 2003, p. 288). A content analysis of public documents is useful for carrying out an exploratory or descriptive study; despite the prevalence of themes that might be shown, however, researchers should not assume that causality has been established. This challenge is not easily

mitigated; therefore, researchers should not use content analysis if the aim of their study is to test a causal relationship.

Overall

Overall, qualitative research relies upon the researchers' examination of naturalistic environments in order to understand the context in which accounting phenomena occur. This presents the researchers with flexibility and enhanced external validity, but must be balanced against researcher biases and the inability to test causal relationships. When used in conjunction with archival and behavioral methods, qualitative methods can provide a richer picture of the setting's context and environment.

EARNINGS MANAGEMENT: TRIANGULATION OF RESEARCH METHODS

Archival, behavioral, and qualitative research methods all have limitations in their ability to explore research questions. Archival research methods are useful in explaining general large-scale phenomena, behavioral methods are well suited for testing causal inferences, and qualitative methods afford a deeper understanding of context. Thus, the limitations inherent in a particular method may be addressed by using one or more complementary methods. For instance, qualitative and archival research methods cannot adequately assess causal relationships, whereas behavioral experiments can, due to the higher internal validity afforded by this method. Behavioral experiments may have limited external validity, however, while archival studies are useful for examining naturally occurring phenomena and qualitative research methods afford a rich examination of context. Hence, using multiple methods to examine common research questions is more robust than a single method in isolation.

The lack of superiority of a single research method points to the need for *triangulation*. Simply stated, triangulation of research methods involves the use of multiple research methods to investigate a research question.³ Triangulation can help researchers improve the *generalizability* of findings, while presenting a *richer* picture of a particular phenomenon. Several researchers have called for the joint use of in-depth field studies with cross-sectional surveys to improve both theory and empirical testing (e.g., [Arnold, 2006](#)). Such a multi-method approach – particularly a mixture of qualitative

and quantitative methods – helps to overcome the weaknesses of using a single research method in isolation (Kerlinger & Lee, 2000). This could best be accomplished by fostering a dialogue across different studies, such that synthesizing multiple papers relying on multiple methods to investigate a common phenomenon could provide a richer, panoramic view of the topic in question.

The advantages of triangulation are evident if researchers' goals are to understand a broad phenomenon. One area of research in accounting, *earnings management*, provides an interesting lens through which to view the usefulness of triangulation. Earnings management is defined by Schipper (1989) as “purposeful intervention in the external reporting process, with the intent of obtaining some private gain to managers or shareholders.” Most of the archival literature on earnings management has sought to *identify* instances of earnings management, and ultimately to *explain under what conditions* firms manage their financial earnings (Koonce & Mercer, 2005). Overall, archival studies on the earnings management phenomenon have failed to provide convincing evidence that managers explicitly misstate earnings. Part of the reason for this lack of support is due to the inherent limitations of archival research. Since archival methods use secondary databases (none of which contain a convenient “earnings management” dataset), researchers have had to identify proxies to test for instances of earnings management. While a variety of models have been developed to proxy for earnings management, all of these demonstrate low power and an inability to properly distinguish between true performance and opportunistic behavior (DeChow, Sloan, & Sweeney, 1995). Some newer models have made improvements (e.g., Francis, LaFond, Olsson, & Schipper, 2005), but archival researchers are faced with an uphill battle in solely using secondary data to study earnings management if they cannot determine when such behavior exists.

A second limitation of archival studies of earnings management is that it has been very difficult to tap into the *motivation* involved with this behavior (Schipper, 1989). Archival studies have used economics-based theory to provide a number of explanations for why earnings management may occur, including political reasons (e.g., Jones, 1991), bonuses (e.g., Watts & Zimmerman, 1986), satisfying debt covenants (e.g., DeChow, Sloan, & Sweeney, 1996), or optimally timing equity transactions (e.g., Kothari, 2001). However, since archival studies cannot easily establish causal inferences, it is difficult to conclude whether firms with these characteristics engage in earnings management because of the above-mentioned motivations, or whether another correlated omitted variable is responsible for the

connection. Likewise, the motivations used in archival studies of earnings management ignore the nuances of human behavior that may also be responsible for such actions (Koonce & Mercer, 2005).

Behavioral studies of earnings management have complemented the findings in the archival literature by addressing some of these shortcomings and provide evidence that earnings management behavior is very prevalent. Whereas archival studies have failed to convincingly demonstrate that managers overtly manage earnings, behavioral studies' use of individual participants provides stronger support for the occurrence of this behavior. Synthesizing the results of archival and behavioral studies can therefore help overcome the limitations of relying upon studies of a single method. Part of the reason for this difference in findings in the archival and behavioral research method streams is due to the varying strengths of these two methods. From a behavioral perspective, "earnings management" behavior has stronger construct validity (i.e., is more easily defined and measurable) than from an archival standpoint. However, behavioral methods are not immune from threats to construct validity. In laboratory studies, researchers must design a task for participants to complete, and the operationalization of "earnings management behavior" may prove similarly challenging.

Behavioral studies on earnings management have often used *auditors* as participants to help gather information about this phenomenon. One such study is Nelson, Elliott, and Tarpley (2002, p. 184), in which the researchers surveyed auditor partners and managers about "experiences with companies that attempt material earnings management." This study provided evidence on the effect of the preciseness of accounting standards on earnings management behavior among both managers and auditors. Specifically, auditors identified that managers were most likely to attempt earnings management behavior by either structuring transactions to satisfy very precise accounting standards (e.g., reserves), or by leaving transactions unstructured to satisfy imprecise standards (e.g., leases); auditors reported they were also less likely to require financial statement adjustments of such behavior. The behavioral method in Nelson et al. (2002) results in stronger construct validity than would attempts to develop an empirical proxy of earnings management, which could be why the study provided stronger evidence of earnings management behavior than seen in the archival literature.

Another manner in which behavioral studies can complement the findings of archival studies is through the strong internal validity of a laboratory experiment, which makes this a very useful method for establishing causal inferences. For instance, Lee, Petroni, and Shen (2006) conducted an archival study that found that firms with a history of earnings management

behavior tended to use less transparent reporting. [Hunton et al. \(2006\)](#) designed an experiment to test the causality of this relationship, and showed that greater transparency of standards helped to mitigate earnings management behavior. These findings demonstrate that archival studies are useful for establishing an overall pattern of behavior among naturally occurring events, whereas behavioral studies can provide evidence on the causality of relationships and are hence able to demonstrate the overt nature of earnings management behavior. Another laboratory experiment in this vein is [Beeler and Hunton \(2002\)](#), which used audit partners as participants and found that auditor judgment was biased by the presence of contingent economic rents. Extrapolating these findings suggests that earnings management behavior may be more common in these circumstances; again, such a causal relationship could not be gleaned from an archival study.

Overall, evidence from behavioral studies demonstrates that purposeful earnings management behavior is indeed common. In particular, evidence from auditors suggests that managers are more likely to manage earnings given fewer opportunities for detection and external constraints. The weaker construct and internal validity found in archival studies is the likely reason why results from behavioral studies provide stronger evidence of earnings management behavior. Thus, utilizing behavioral methods in conjunction with archival research affords a deeper understanding of the topic of earnings management.

Despite the broadened view given by a joint examination of a topic from archival and behavioral viewpoints, both of these methods ignore the *context* within such behavior occurs. The use of a qualitative field study to address this topic could help researchers “drill deeper” into the underlying rationale behind earnings management and provide a better understanding of the findings from the archival and behavioral studies thus far. In the words of [Merchant and Van der Stede \(2006, p. 129\)](#), “... [I]f more researchers would engage in field research and ask management how and why they would manage earnings ... then a richer theory would emerge with details about the methods used to manage earnings and multiple relevant behavioral factors.” Archival research tends to truncate outliers; a qualitative field study could focus on extremes of behavior and help to assess how corporate environments and/or individual management factors interacted to exacerbate or constrain earnings management behavior ([Merchant & Van der Stede, 2006](#)). Such a qualitative study could involve interviewing managers to determine how they defined and recognized “earnings management,” along with ascertaining its contributing factors. This would help to tie together past archival and behavioral studies on earnings management and develop a more complete picture of the process.

Overall, triangulating research methods in investigating a complex accounting phenomenon such as earnings management affords a richer, panoramic view of the motivations, causality, context, and contributing factors of this occurrence. Despite these potential benefits, however, many issues may arise when researchers attempt to synthesize separate studies that have used differing research methods to investigate a common phenomenon. Two of the most salient issues concern the role of research methods in different stages of investigation, and the reconciliation of conflicting results.

First, an unanswered question is whether all research methods are appropriate for all stages of investigating a phenomenon, or whether specific methods have a distinct role at different phases. For example, in the study of earnings management, archival methods may be useful for investigating patterns of behavior associated with the presence of earnings management, behavioral methods can follow up with more rigorous tests of causality of the purported relationships, and qualitative methods can build upon such work by delving in to more complex “how” and “why” relationships. However, these methods may still have a role at other stages – for example, qualitative methods may be useful in the initial stages of investigation in helping to define a phenomenon’s constructs. Researchers should therefore think carefully about the aims of their study before deciding on a research method.

Second, using multiple research methods across research studies opens the door to the possibility of conflicting results. For example, a behavioral experiment that follows up on an archival study may fail to find the hypothesized causal claim between constructs. Moreover, qualitative field studies may prove that the construct is more complex than initially anticipated, suggesting that prior proxies used to operationalize this construct may have been poor choices. While there is no clear answer for reconciliation, the conflicting results achieved with multiple research methods could suggest that further research be conducted on the phenomenon – perhaps by re-examining the theory suggesting causality, or by developing better proxies. In all, triangulation of multiple research methods does not guarantee a perfectly aligned picture of a phenomenon, but could lead to advances in incremental knowledge.

CONCLUSION

This chapter presents an overview of the limitations of relying on research that exclusively uses archival, behavioral, or qualitative research methods. Triangulation of studies that have used differing research methods will help

researchers overcome the limitations of reliance on a single method and achieve a deeper understanding of accounting phenomena. In particular, recognizing the oft-neglected value of behavioral and qualitative research would be beneficial to the academic accounting community, particularly as such methods may help compensate for some of the inherent weaknesses in archival research methods.

Triangulation of multiple research methods has the potential to increase both the generalizability and richness of findings for a given phenomenon. Research methods are the tools researchers utilize to test research questions; the joint use and understanding of multiple types of research methods affords a broader view of appropriate research questions and of accounting research in general. Some evidence suggests that researchers are beginning to realize the benefits of triangulation. For example, the use of an experiential questionnaire enables practitioners to provide in-depth information about real-world phenomena within a structured environment; this method thus provides the richness of an interview (qualitative) with the structure of a questionnaire (behavioral) (Gibbins & Qu, 2005). Unfortunately, the use of multiple research methods is rare, as are attempts to connect studies of the same topic that were investigated with different research methods (Lillis & Mundy, 2005; Merchant & Van der Stede, 2006). This suggests that the accounting research community has much to gain by broadening its collective paradigmatic lenses (e.g., Lewis & Grimes, 1999).

Triangulating across methods and conducting multi-method research are not without their own limitations. A broader view of research entails much larger time and resource commitments for researchers to become familiar with studies using different methods, not to mention the huge start-up costs researchers must invest in learning how to use different research methods themselves. Due to the doctoral education process, much research is conducted in “silos” that may be difficult to eradicate overnight (Koonce & Mercer, 2005). Moreover, triangulating studies that have used differing research methods to examine the same research question may produce conflicting results that point the way for further work. Nevertheless, being aware of multiple methods and recognizing the value and role of each are important early steps in broadening the research process. This awareness is critical so that researchers can look beyond their “native” research stream in helping to motivate their own studies. While it may not be feasible for researchers to conduct research using different types of methods, they should be open to incorporate the results gained from other research methods in order to better understand the holistic nature of a research phenomenon.

In conclusion, accounting is essentially an inter-disciplinary area formed at the crossroads of several root disciplines and research methods. To borrow from Slemrod (2003), an attempt to study accounting phenomenon through the singular lens of a particular research method is like the parable of the elephant and the blind men – each are incapable of recognizing that they are faced with an elephant due to their preoccupation with a narrowly defined realm. By breaking down the barriers between research methods, accounting researchers can attempt to examine the “larger creature” behind accounting phenomena.

NOTES

1. In trigonometry and geometry, the term *triangulation* refers to the process of locating the coordinate of and distance to a target point on a plane, given the distances and angles from two reference points. The process of triangulation involves the law of sines and the Pythagorean theorem. The concept of triangulation has been adopted by researchers to mean that if researchers use, say, two different methods to study a phenomenon and both methods generally arrive at the same result, then they are in a much better position to draw a generalizable inference, relative to the use of a single method. Although the mathematical concept of triangulation used in trigonometry and Euclidean geometry involves three points (two known and one unknown), researchers have broadened the concept to include two or more different methods.

2. While archival research examines data originating from a wide cross-section of firms or other units, the majority of accounting studies employing this method tend to rely on large secondary databases with more observations than are typically seen in behavioral or qualitative studies.

3. The goal of “triangulation” is not to use multiple methods within the same study, but rather to recognize that separate papers relying on differing methods that investigate a common phenomenon all have value in helping to develop a comprehensive view of the research phenomenon.

REFERENCES

- Anderson-Gough, F., Grey, C., & Robson, K. (2005). ‘Helping them to forget...’: The organizational embedding of gender relations in public audit firms. *Accounting, Organizations and Society*, 30, 469–490.
- Annisette, M. (2000). Imperialism and the professions: The education and certification of accountants in Trinidad and Tobago. *Accounting, Organizations and Society*, 25, 631–659.
- Arnold, V. (2006). Behavioral research opportunities: Understanding the impact of enterprise systems. *International Journal of Accounting Information Systems*, 7, 1–11.

- Arnold, V., Collier, P. A., Leech, S. A., & Sutton, S. G. (2000). The effect of experience and complexity on order and recency bias in decision making by professional accountants. *Accounting and Finance*, 40, 109–134.
- Ashton, R. B. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, 28(Suppl.), 148–180.
- Bailey, C., Hasselback, J., & Karcher, J. (2001). Research misconduct in accounting literature: A survey of the most prolific researchers' actions and beliefs. *Abacus*, 37, 26–54.
- Baker, R. (2005). What is the meaning of "the public interest"? Examining the ideology of the American public accounting profession. *Accounting, Auditing and Accountability Journal*, 18, 690–703.
- Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 7, 159–178.
- Beaver, W. (1968). The information content of annual earnings announcements. *Journal of Accounting Research* (Suppl.), 67–92.
- Beeler, J. D., & Hunton, J. E. (2002). Contingent economic rents: Insidious threats to audit independence. *Advances in Accounting Behavioral Research*, 5, 21–50.
- Berg, B. L. (2003). *Qualitative research methods for the social sciences*. New York: Allyn and Bacon.
- Bobek, D., & Radtke, R. (2007). An experiential investigation of tax professionals' ethical environments. *Journal of the American Taxation Association*, 29, 84.
- Bonner, S., & Sprinkle, G. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence and a framework for research. *Accounting, Organizations and Society*, 27, 303–345.
- Bonner, S. E., & Walker, P. L. (1994). The effects of instruction and experience on the acquisition of auditing knowledge. *The Accounting Review*, 69, 157–178.
- Broadbent, J., & Laughlin, R. (1997). Developing empirical research: An example informed by a Habermasian approach. *Accounting, Auditing and Accountability Journal*, 10, 622–648.
- Broadbent, J., & Laughlin, R. (1998). Resisting the 'new public management': Absorption and absorbing groups in schools and GP practices in the UK. *Accounting, Auditing and Accountability Journal*, 11, 403–435.
- Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organizational analysis*. London: Erlbaum.
- Canada, J., Kuhn, J. R., & Sutton, S. G. (2008). *The pervasive nature of IT controls: An examination of material weaknesses in IT controls and audit fees*. Working paper. University of Central Florida.
- Chua, W. F., & Poullaos, C. (2002). The Empire strikes back? An exploration of centre-periphery interaction between the ICAEW and accounting associations in the self-governing colonies of Australia, Canada, and South Africa, 1880–1907. *Accounting, Organizations and Society*, 27, 409–445.
- DeChow, R. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *The Accounting Review*, 70, 193–225.
- DeChow, R. M., Sloan, R. G., & Sweeney, A. P. (1996). Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research*, 13, 1–36.
- Dillman, D. A. (2006). *Mail and Internet surveys: The tailored design method*. New York: Wiley.

- Elliott, W. B., Hodge, F. D., Kennedy, J. J., & Pronk, M. (2007). Are MBA students a good proxy for non-professional investors? *The Accounting Review*, 82, 139–168.
- Francis, J., LaFond, R., Olsson, P., & Schipper, K. (2005). The market pricing of accruals quality. *Journal of Accounting and Economics*, 39, 295–327.
- Gibbins, M. (1992). Deception: A tricky issue for behavioral research in accounting and auditing. *Auditing: A Journal of Practice and Theory*, 11, 113–126.
- Gibbins, M., & Qu, S. Q. (2005). Eliciting experts' context knowledge with theory-based experiential questionnaires. *Behavioral Research in Accounting*, 17, 71–88.
- Gramling, A. A., Johnstone, K. M., & Mayhew, B. W. (2001). Behavioral research in auditing: Past, present, and future research. *Advances in Accounting Behavioral Research*, 4, 47–75.
- Greene, W. H. (2003). *Econometric analysis*. Hillsdale, NJ: Pearson Education.
- Greenwood, R., Suddaby, R., & Hinings, C. R. (2002). Theorizing change: The role of professional associations in the transformation of institutionalized fields. *Academy of Management Journal*, 45, 58–80.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Hillsdale, NJ: Pearson Education.
- Hampton, C. (2005). Determinants of reliance: An empirical test of the theory of technology dominance. *International Journal of Accounting Information Systems*, 6, 217–240.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1–55.
- Holthausen, R., Larcker, D., & Sloan, R. (1995). Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics*, 22, 29–74.
- Hunton, J. E., Libby, R., & Mazza, C. L. (2006). Financial reporting transparency and earnings management. *The Accounting Review*, 81, 135–157.
- Jones, J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, 29, 193–228.
- Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, 31, 231–245.
- Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. New York: Thomson Learning.
- Klein, A. (2002). Audit committee, board of director characteristics, and earnings management. *Journal of Accounting and Economics*, 33, 375–400.
- Koonce, L., & Mercer, M. (2005). Using psychology theories in archival financial accounting research. *Journal of Accounting Literature*, 24, 175–214.
- Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, 31, 105–231.
- Lee, Y.-J., Petroni, K. R., & Shen, M. (2006). Cherry picking, disclosure quality, and comprehensive income reporting choices: The case of property-liability insurers. *Contemporary Accounting Research*, 23, 655–700.
- Leech, S. A., Collier, P. A., & Clark, N. (1998). Modeling expertise: A case study using multiple corporate recovery experts. *Advances in Accounting Information Systems*, 6, 85–105.
- Lewis, M. W., & Grimes, A. J. (1999). Metatriangulation: Building theory from multiple paradigms. *Academy of Management Review*, 24, 672–690.
- Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society*, 27, 775–810.

- Lillis, A. M., & Mundy, J. (2005). Cross-sectional field studies in management accounting research: Closing the gaps between surveys and case studies. *Journal of Management Accounting Research*, 17, 119–141.
- Llewelyn, S. (2003). What counts as “theory” in qualitative management and accounting research? Introducing five levels of theorizing. *Accounting, Auditing, and Accountability Journal*, 16, 662–708.
- Merchant, K. A., & Van der Stede, W. M. (2006). Field-based research in accounting: Accomplishments and prospects. *Behavioral Research in Accounting*, 18, 117–134.
- Mouritsen, J., & Thrane, S. (2005). Accounting, network complementarities and the development of inter-organizational relations. *Accounting, Organizations and Society*, 30, 1–35.
- Nelson, M. W., Elliott, J. A., & Tarpley, R. L. (2002). Evidence from auditors about managers’ and auditors’ earnings management decisions. *The Accounting Review*, 77, 175–202.
- Nelson, M. W., Libby, R., & Bonner, S. E. (1995). Knowledge structure and the estimation of conditional probabilities in audit planning. *The Accounting Review*, 70, 27–47.
- Nicolaou, A. I. (2004). Quality of post-implementation review for enterprise resource planning systems. *International Journal of Accounting Information Systems*, 5, 25–49.
- Omer, T. C., & Shelley, M. K. (2004). Competitive, political, and economic factors influencing state tax policy changes. *Journal of the American Taxation Association*, 26(Suppl.), 103–126.
- O’Brien, P. (1988). Analysts’ forecasts as earnings expectations. *Journal of Accounting and Economics*, 10, 53–83.
- O’Neil, C. J., & Samelson, D. P. (2001). Behavioral research in taxation: Recent advances and future prospects. *Advances in Accounting Behavioral Research*, 4, 103–139.
- Schipper, K. (1989). Commentary on earnings management. *Accounting Horizons*, 3, 91–102.
- Seal, W., Berry, A., & Cullen, J. (2004). Disembedding the supply chain: Institutionalized reflexivity and inter-firm accounting. *Accounting, Organizations and Society*, 29, 73–92.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.
- Simunic, D. (1980). The pricing of audit services: Theory and evidence. *Journal of Accounting Research*, 18, 161–190.
- Slemrod, J. (2003). Tax from any angle: Reflections on multi-disciplinary taxation research. *National Tax Journal*, 56, 145–151.
- Sprinkle, G. B. (2000). The effect of incentive contracts on learning and performance. *The Accounting Review*, 75, 299–326.
- Tubbs, R. M. (1992). The effect of experience on the auditors’ organization and amount of knowledge. *The Accounting Review*, 67, 783–801.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- Watts, R., & Zimmerman, J. (1986). *Positive accounting theory*. Englewood Cliffs, NJ: Prentice Hall.

DISCLOSURE VERSUS RECOGNITION IN STOCK-OPTION REPORTING: ARE SOPHISTICATED USERS' PERCEPTIONS AND JUDGMENTS INFLUENCED BY THE REPORTING FORMAT?

Asokan Anandarajan, Réjean Belzile,
Anthony P. Curatola and Chantal Viger

ABSTRACT

The recently passed Statement Financial of Accounting Standard (SFAS) 123R mandates that stock-option compensation costs be recognized in the income statement. This supersedes SFAS 148 and the earlier SFAS 123 which required only disclosure in the notes to the financial statements. The motivation of the Financial Accounting Standards Board (FASB) was to increase transparency in reporting of financial statements. The objective of this chapter is to test whether sophisticated users' perceptions and judgments are affected by the different reporting format that has been mandated by SFAS 123R. Members of the Institute of Management Accountants (IMA) were used

Advances in Accounting Behavioral Research, Volume 11, 31–57

Copyright © 2008 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11002-X

as the participants in this study. The study finds a (1) higher perceived risk, (2) lower expected accounting return, (3) more pessimistic overall perception, (4) more negative future stock price direction, and (5) lower stock price valuation by sophisticated users in the presence of recognition versus disclosure. These findings support the stance of the FASB and indicate that that information content is accentuated in the presence of recognition relative to disclosure.

INTRODUCTION

In December 2004, the Financial Accounting Standards Board (FASB) issued Statement of Financial Accounting Standard (SFAS) 123R requiring that companies recognize the expense of stock options in the income statement. Prior to this on March 31, 2004, FASB issued an exposure draft entitled *Accounting for Share Based Payments*, which mandated the recognition of the expense associated with employee stock options; and, it was met with considerable controversy. In fact, there was an attempt in Congress to have this rescinded. The effort, however, failed due to a lack of support in the Senate.

The motivation for SFAS 123R is aptly summarized by Robert Herz, Chairman of the FASB testifying before a subcommittee in Congress on April 20, 2004, in which he said:

Recent events have served as a reminder to all of us that clear credible and comparable financial information is essential to the health and vitality of our capital market system. In the wake of the market meltdown and corporate reporting scandals, the FASB has received numerous requests from individual and institutional investors, financial analysts and many others urging the Board to mandate the expensing of the compensation cost relating to employee stock options. (Herz, 2004)

Herz also noted that recognition, as opposed to disclosure of stock-option costs and its impact on profits in the notes to the financial statements (hereafter referred to as disclosure), provides greater transparency to financial statement users, hence the demand for recognition. A move to require expensing can be seen as another step toward the convergence of United States and international accounting standards.¹ The objective of this study is two-fold. Using the FASB standards relating to stock options, the objective is (1) to examine whether recognition conveys a stronger signal

relative to disclosure in the notes and (2) to provide further evidence for the FASB's stance regarding expensing of stock options.

To examine this issue, members of the Institute of Management Accountants (IMA) were solicited to complete one of two case studies. One case provided information regarding stock-option costs with an attached pro forma statement in the notes to the financial statements. The other case recognized the stock-option costs in the company's income statement. Participants were asked to indicate their assessment of the investment risk, the expected accounting return, and the overall risk associated with this hypothetical company. Participants were also asked to indicate their judgments on (1) the expected future stock price direction and (2) the stock price value.

The results indicate that sophisticated users estimate a higher risk, lower expected accounting return, and perceive a higher overall risk when stock-option costs are recognized. They also estimate a more pessimistic expected future stock price direction and a lower stock value in the presence of recognition relative to disclosure. While participants find that the financial statements are more helpful in the presence of recognition versus disclosure, they do not consider recognition to make the financial statements significantly less confusing, or to significantly increase their reliability and clarity.

This study contributes to the literature by providing insights into the potential consequences of SFAS 123R. These findings suggest that the stronger signal conveyed by recognition relative to disclosure influences sophisticated investors' perceptions and judgments. Thus, the study's findings corroborate the stance of the members of the FASB who mandated recognition on the presumption that recognition enhances transparency relative to disclosure. Likewise the findings are consistent with views expressed by the Financial Accounting Standards Committee of the American Accounting Association (hereafter the AAA Committee) which is charged with responding to requests for comment from standard setters on issues relating to financial reporting. Recently, the [American Accounting Association Financial Accounting Standards Committee](#) commented in *Accounting Horizons* that "the Committee strongly endorses the conclusion that share-based payments should lead to expense recognition" (p. 103). They also noted that "disclosure is not an adequate substitute for recognition" (p. 104). These findings also corroborate the results of two newly published studies reporting that recognition relative to disclosing stock options influences non-sophisticated investors' judgments (Belzile, Fortin, & Viger, 2006) and loan officers' judgments and decisions (Viger, Belzile, & Anandarajan, 2008).

BACKGROUND AND LITERATURE REVIEW

Accounting for stock options historically has been viewed as a form of non-accounting because companies were allowed to blur the potential bottom line impact of stock options in a footnote to their financial statements. Critics suggest that this form of reporting ignores economic reality and is inconsistent with accounting standards for other types of compensation. Regulations in stock-option reporting date back to 1972. In that year Accounting Principle Board (APB) Statement No. 25 acknowledged that, because of measurement considerations,² it did not require accrual of compensation cost. By disregarding the option's value, APB 25 ignored the fact that options give employees valuable rights and that options are compensation for services performed.

In 1995 FASB issued SFAS No. 123, which mandated footnote disclosure of pro forma net income and EPS as if the expense (using the fair value method) had been incorporated in the income statement if the intrinsic value was selected for the measurement at the granting date.³ The intent of adding the pro forma figures was to provide a clearer signal to financial statement readers about the potential financial impact of stock-option expenses. SFAS 123 also encouraged firms to report expenses in the income statement using the fair value method. Not surprisingly, few companies elected to use the fair value method until the recent scandals of Enron and WorldCom. A few firms such as DuPont, Conoco Philips, Gabelli Asset Management, Microsoft Corporation, The Washington Post Company, and Coca Cola Corporation voluntarily began expensing stock-option costs in 2003. The majority of companies, however, continued to show these costs as a footnote disclosure to their financial statements.

To provide guidance for the transition from pro forma footnote disclosure to mandatory recognition in the income statement, FASB issued SFAS 148 entitled *Accounting for Stock Based Compensation-Transition and Disclosure* in December 2002. However, it did not mandate recognition of expenses in the income statement. On March 31, 2004, FASB issued an exposure draft entitled *Accounting for Share Based Payments*, which mandated the recognition of the expense associated with employee stock options at fair value. The motivation of the FASB was to ensure convergence with international accounting standards which would require expensing of stock-option costs of fiscal periods beginning on or after January 1, 2005. The House of Representative, sought to repeal this through the passage of the Stock Option Reform Act (HR 3574) in May 2004. This Act basically espoused continuing the current treatment of footnote

disclosure of stock-option costs. The Senate did not show the same level of interest as the House on this issue. In fact, this legislation caused controversy and contentious debate in the Senate. Representative Pete Stark (Democrat California) noted “let’s help the rich get richer. Right now, companies can deduct the cost of stock options for income tax purposes, but don’t have to report that value to shareholders.”

Despite the contention that arose, the current standard, SFAS 123R, requires mandatory expensing of stock-option costs. Public firms are now required to estimate the fair value of stock options at the granting date and then expense it (between the granting date and the earliest exercisable date) from revenues. In essence, all forms of share-based payment to employees, including employee stock options, are now treated the same as other forms of compensation by recognizing the related cost in the income statement.

Empirical research suggests that placement of the stock-option expense within the financial statements, whether in the income statement or only in the footnotes, makes *little difference* to investors (Balsam, Bartov, & Yin, 2004). Prior studies examining stock price reaction to footnote disclosure regarding pensions and post-retirement benefits, however, found that footnote disclosures do have informational content (Barth, 1991; Choi, Collins, & Johnson, 1997). These findings corroborate the efficient market hypothesis, which implies that once information is disclosed investors knowingly incorporate that information into their investment decision-making.

Initial studies on the topic of interest in this chapter, namely presentation of stock-option information, investigated the relationship between SFAS 123 annual option values and stock prices. These studies sought to test whether the pro forma disclosure mandated by SFAS 123 had information content. The studies hypothesized that the information would convey a significant negative signal to the market. Bell, Landsman, Miller, and Yeh (2002) investigated a sample of 85 profitable companies from the software industry in the 3-year period of 1996–1998. They failed to find their expected negative relation between annual stock-option expense disclosed in the footnotes and stock prices. Aboody, Barth, and Kasznik (2004) used the same 3-year sample period with 2,274 firm years and documented a negative correlation between disclosed annual option expense and stock prices. They concluded that pro forma disclosure does convey information to investors. Aboody et al. (2004) explained the differences in results by concluding that the Bell et al. (2002) findings were sample specific and could not be generalized to other samples.

In another recent study [Balsam et al. \(2004\)](#) examined the same relationship for firms for the first year SFAS 148 became effective by examining the information content of both recognition and pro forma footnote disclosure. They found that *placement* of the stock-option information (whether on the income statement or only in the footnotes as pro forma disclosure) made little difference to investors. In their study, the cumulative abnormal return were regressed on recognized option expense and unrecognized option expense after controlling for variables shown by prior research to explain stock returns. Their results indicate that, while both forms of disclosure are value relevant, there is no significant difference in market reaction between them. They concluded that the economic consequences of option expense recognition are unlikely to be significant thereby addressing a major concern of opponents of recognition of stock-option costs in the income statement. Despite the findings of research that indicate users consider information provided in financial statement footnotes, the AAA Committee supported the position held by the FASB that disclosure is not an adequate substitute for recognition.

From an experimental study point of view, the communication paradigm suggests that location of information does have an influence on financial statements users' perceptions and decisions ([Viger et al., 2008](#)). [Hirst and Hopkins \(1998\)](#) found that professional analysts are more likely to discover earnings management when earnings components are clearly reported in a performance statement than when they are disclosed in notes that require more analysis. Further, [Hirst, Hopkins, and Wahlen \(2004\)](#) examined the influence of footnote disclosure of fair value instruments and concluded that disclosure was *not a substitute* for full fair value income measurement. This was also found to be true with respect to placement and location of going concern information ([Viger, Anandarajan, Curatola, & Amar, 2004](#); [Anandarajan, Viger, & Curatola, 2002](#) among others) and disclosure of comprehensive income ([Hirst et al., 2004](#) among others). Two recent experimental studies conducted in the context of recognition/disclosure of stock-option compensation conclude that presentation format influences non-professional investors ([Belzile et al., 2006](#)) and loan officers ([Viger et al., 2008](#)).

In contrast, [Frederickson and Miller \(2004\)](#) examined the effect of pro forma earnings disclosures on investors' judgments. They concluded that, while non-professional (less sophisticated) investors were misled by favorable non-GAAP pro forma disclosures, analysts (sophisticated investors) acquired the information and were shrewd enough not to be deceived by the same information. Thus, the debate with respect to

professional users is still unresolved as revealed by two recent studies (Hirst et al., 2004; Frederickson & Miller, 2004). Assuming that the location and placement of information does influence financial statement users, a significant difference between pro forma disclosure and recognition on investors' judgments and valuations should be expected.

Empirical studies, as noted above, focus on the reaction of the broader market; therefore, their results do not apply to individual investors. Experimental studies such as this, add to the extant literature by examining whether sophisticated individual investors are influenced by the two methods of reporting stock options (recognition versus disclosure). Most studies that used sophisticated users have investigated their job-related judgments; this study focuses on the judgments of sophisticated users in the context of their personal investment decision-making. This study is different from prior studies because we use members of the IMA as opposed to MBA students as surrogates for sophisticated investors.

HYPOTHESES

Financial statement users (e.g., investors) are perceived to be imperfect processors of information (Hirshleifer & Teoh, 2003). Assuming imperfection, Fiske and Taylor (1991) note that the process of interpreting information (which they call "coding") involves taking external information and assembling it in a meaningful way internally. The vast amount of information, however, presented in financial statements and elsewhere requires financial statement users to be selective with respect to information cues (Kahneman, 1973; Kahneman & Tversky, 1973). That is because there are limits to information processing. Libby, Bloomfield, and Nelson, (2002) and Maines (1995) note that placement, categorization, and labeling all play a role in this form of simplification and even influence professional financial statement users as they evaluate accounting information. Placement, categorization, and labeling enable some information to be perceived and encoded more easily than others. Hirshleifer and Teoh (2003) call this concept *saliency*, i.e., a signal's ability to stand out. Saliency, in turn, influences judgment. Hirshleifer and Teoh (2003) note that if saliency of a disclosure is not high, some financial statement users may fail to process it.

Other causes have been suggested by previous research for the faulty acquisition and/or processing of financial information. Users limit the time and effort they put into acquiring and analyzing accounting data

(Bloomfield, 2002; Barberis & Thaler, 2003) or are fixated on reported earning numbers (Ijiri, Jaedicke, & Knight, 1966; Tinic, 1990). According to a signaling explanation, management may attempt to signal that information is not dependable by disclosing instead of recognizing it (Bernard & Schipper, 1994), inducing users to disregard it. An immediate but far-reaching consequence is that reporting can have different effects on the perceptions, judgments, and decisions of financial statement users, depending on the form of presentation.

Assuming that sophisticated investors' perceptions and judgments are influenced by the recognition of the expense in the financial statements (rather than disclosure of the information in the footnotes), we expect that lower reported earnings under recognition will give rise to a less favorable overall judgment of the company. Hence, the first hypothesis is stated as follows:

H₁. Individual's overall underlying judgment of the company will be more pessimistic when stock-option costs are recognized in the income statement than when they are disclosed in a footnote.

Aboddy et al. (2004) in an empirical study find that share prices are negatively associated with stock-based compensation recognized in the income statement. They conclude that recognition may convey an impression of weaker financial results to investors relative to disclosure. Similarly, Hirshleifer and Teoh (2003) indicate that inattentive investors extrapolate reported earnings to formulate an opinion on a company. The implication is that recognition of stock-option compensation may have an adverse impact on expected stock price. Based on the theory espoused by Hirshleifer and Teoh (2003) and the findings of Aboddy et al. (2004), recognition should have a more adverse impact on perceived stock price value relative to disclosure. Our second and third hypotheses are stated as follows:

H₂. Investors will predict a more pessimistic expected future stock price direction when stock-option costs are recognized in the income statement than when they are disclosed in a footnote.

H₃. Investors will estimate a lower stock price value when stock-option costs are recognized in the income statement than when they are disclosed in a footnote.

METHODOLOGY

Sample Selection

Participants for this study were randomly drawn from a membership list of the IMA. Members of the IMA represent a diverse mix of investors. They are heterogeneous and perform a wide range and variety of tasks in a variety of accounting related jobs. This makes this population particularly appealing for this task; they are qualified to critically analyze stock-option information and as such are deemed sophisticated users.

The data was collected for this study in two forms. First, a case study and survey instrument was initially provided to participants on a website. The sample for the website participation was a random selection made by the headquarters of IMA. The members were requested to enter the website, download the material, and after studying the case study, fill out the survey instrument. An e-mail request was sent to a randomly selected group of 4,000 IMA members to voluntarily participate in our study. A total of 42 completed and usable responses were downloaded from the website. Second, IMA members attending an IMA meeting were requested to participate in this study. A hard copy of the case was provided to and collected from 26 participants in this group. Our final sample consisted of 68 participants. A test comparing the demographic variables revealed no significant difference between the two experimental groups.

Research Design

The research design is shown in [Fig. 1](#). Each participant received an experimental package that included descriptive information on a hypothetical company. The financial statements included a standard (unqualified) auditor's report, a set of financial statements for two fiscal periods including a balance sheet, income statement, statement of retained earnings, and statement of cash flow and accompanying notes. The information in the two case studies was identical except for the manipulation – the mode of communication of the stock-option compensation. In one case study the stock-option compensation information was reported in the notes to the financial statements (participants receiving this case study are referred to as the disclosure treatment), which also includes pro forma net income and EPS as if the fair value of stock options had been recognized. In the second case study the stock-option

		Experimental groups			
		Group 1 Disclosure treatment group		Group 2 Recognition treatment group	
Experimental material	Each participant received an experimental package that included descriptive information of an hypothetical company (named ABC Inc.), a standard auditor report, a set of financial statements for two fiscal periods including a balance sheet, income statement, retained earnings statement and cash flow statement, and accompanying notes and information on a potential investment decision.				
Treatment: Mode of communication of the stock option compensation	The stock option compensation is reported with a descriptive note including proforma net income		The stock option compensation is reported with a descriptive note and is recognized in the income statement		
Volatility Net income for the year ended November 30	<u>2003</u> \$159,071	<u>2002</u> \$148,497	<u>2003</u> \$41,586	<u>2002</u> \$122,795	

Fig. 1. Experimental Design.

compensation was reported in the notes to the financial statements and was also recognized in the income statement (participants receiving this case study are referred to as the recognition treatment).

Research Instrument

The methodology used in this study is based on prior experimental research that examined information content of various forms of disclosure/reports. In those studies, significant variations in perception of risk and financial health were taken as evidence that one form of disclosure/report format conveyed a “stronger signal” than another (Elias & Johnston, 2001; LaSalle & Anandarajan, 1997; Bamber & Stratton, 1997; Gul, 1987; Bailey, 1981; Libby, 1979; Belzile et al., 2006; Viger et al., 2008 among others). In line with these studies, we asked similar questions (albeit modified for this study).

The response portion of the survey instrument included three sections. Section I asked questions pertaining to participants’ underlying and primary judgments as well as questions related to their perception of quality of the

accounting information provided to them. The underlying judgments (H_1) were measured using questions relating to the perceived risk which consists of the (1) perceived investment risk, (2) overall financial condition, (3) ability to pay its debts, (4) expected accounting return, and (5) the overall perception (namely the perceived investment attractiveness). The expected accounting return consists of the (1) expected return on equity, (2) future net income direction, and (3) growth prospect. The primary judgment (H_2 and H_3) referred to the expected (1) future stock price direction and (2) stock price value.⁴

Sections II and III focused on demographic and manipulation questions. Specifically, participants were asked (1) whether ABC reported pro forma compensation cost figure in the notes, (2) if ABC recorded a stock-option expense in the financial statements, and (3) to recall the net income trend. The final section of the instrument requested standard demographic information about the participants such as their education background, risk tolerance, and gender among others.

Fig. 2 provides a framework to analyze the effects of the reporting format of stock options (recognition or disclosure) on underlying and primary judgments. As shown the reporting format is expected to influence the information processing (which consists of acquisition, evaluation, and weighting of the information as suggested by Maines & McDaniel, 2000). Whether the participants correctly perceive the information is tested with our manipulation check questions described above. Once the information is acquired, the information is then appraised (though the evaluation and weighting steps that are not investigated in this study) which lead to specific performance assessment judgments. We postulate, as shown in Fig. 2, that the underlying judgments (related to the (1) perceived risk, (2) expected accounting return, and (3) overall perception) will influence the primary judgments of the expected (1) future stock price direction and (2) stock price value.

Overall the underlying judgment of the company was examined using a variety of questions relating to (1) perceived risk (which consist of the perceived investment risk, overall financial condition, ability to pay its debts), (2) expected accounting return (which consist of the expected return on equity and future net income direction), and (3) overall perception (namely the perceived investment attractiveness). These questions were used in previous research (Belzile et al., 2006; Viger, et al., 2008; Elias & Johnston, 2001; LaSalle & Anandarajan, 1997; Bamber & Stratton, 1997; Gul, 1987; Bailey, 1981; Libby, 1979).

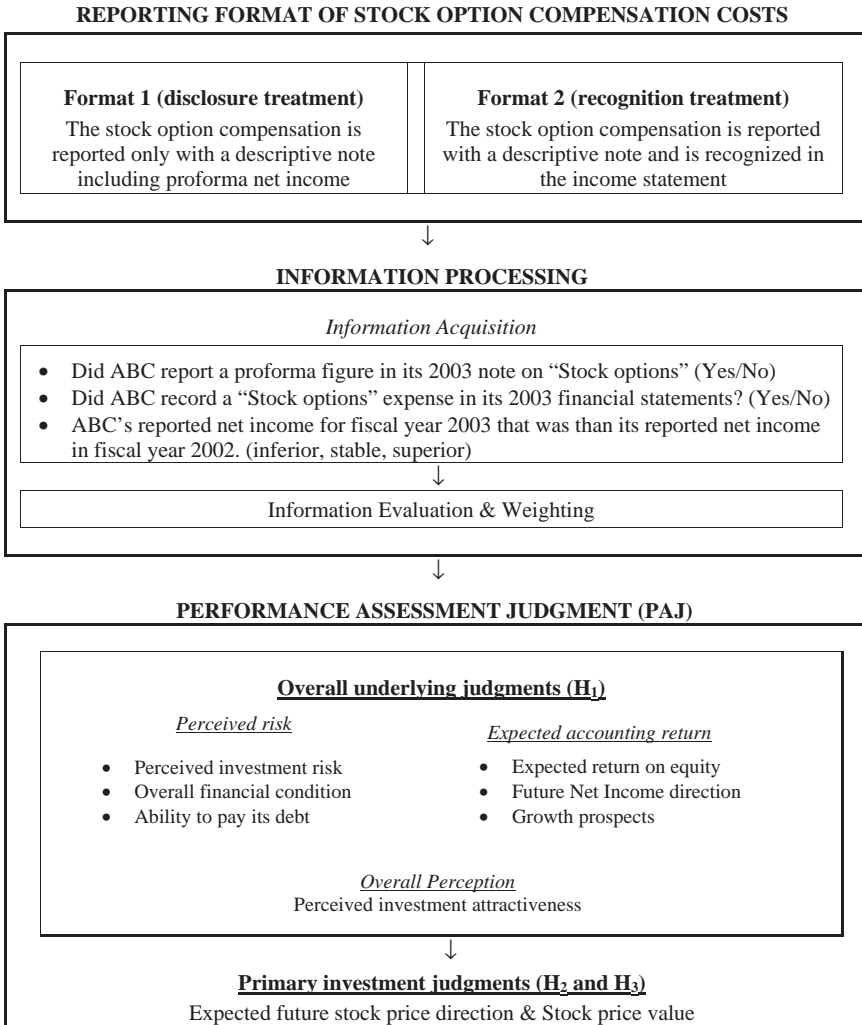


Fig. 2. Framework for the Effects of the Reporting Format of Stock-Option Costs on Underlying and Primary Judgments (Adapted from Maines and McDaniel, 2000).

RESULTS

Preliminary Check

Descriptive statistics on demographic variables (see Table 1) indicate that randomization was successful in obtaining equal distribution among the groups with respect to all demographic variables. The difference between the groups was not statistically significant in terms of educational background, risk tolerance, gender, description of actual portfolio (percentage of companies and in dollar value that is being traded on the NYSE, NASDAQ, and other exchanges), industry in which they tended to avoid investing. Most of the participants had at least general vocational college. With respect to gender, the majority of participants in both groups were males (80% and 70%, in the disclosure treatment and the recognition treatment respectively). Although the participants in the disclosure group had a greater number of participants who categorized themselves as risk seekers (on a Likert scale ranging from 1 (risk averse) to 9 (high risk seeker)), the difference was marginally significant (p -value 0.0935). On an average, participants invested primarily in NYSE (69% for both the disclosure treatment and the recognition treatment in terms of percentage of companies traded in; 65% in the disclosure treatment and nearly 72% in the recognition treatment in terms of dollar value). The results show that 54% in the disclosure treatment and 64% in the recognition treatment indicate that there was no industry they tended to avoid.

Manipulation checks were also conducted. As previously noted, three questions were asked of each participant to check the effectiveness of the manipulation. The results of these three questions are shown in Table 2. Participants were first asked whether ABC reported a pro forma figure in a note in the financial statements. For those participants receiving case one (i.e., the pro forma note disclosure), an affirmative response was given by all participants, which is the correct response. For those participants receiving case two (i.e., the income statement recognition), a negative response was given by 91% of the participants, which is the correct response. Secondly participants were asked whether the hypothetical company recorded stock-option expense in the 2003 financial statements. For the participants receiving case one, the correct response was given by 80% (28 of the 33 participants). For the participants receiving case two, the correct response was given by 82% (27 out of 33 participants). With respect to the third manipulation question, 94% (88%) of the participants in the disclosure treatment (the recognition treatment) correctly responded that 2003

Table 1. Statistics of Demographic Variables: Comparison by Experimental Groups.

	Group	N	Mean	Standard deviation	t	p-value
<i>Educational background</i>						
What is your highest educational background?	1	35	2.24	0.55	-1.61	0.1132
	2	33	2.45	0.56		
(1) High school						
(2) General and vocational college						
(3) Bachelor's degree						
(4) Master's degree						
(5) Doctorate degree						
<i>Risk tolerance (type of investor)</i>						
When it comes to my investment decisions, I generally consider myself _____ (on a 9 points Likert scale where 1 represents risk averse and 9 represents a risk seeker)	1	35	5.94	1.68	1.70	0.0935
	2	33	5.24	1.71		
	Group	N	Male	Female	Pearson χ^2	p-value
<i>Gender</i>						
Male/Female	1	35	80%	20%	0.9616	0.3268
	2	33	70%	30%		

	Group	N	Mean (Standard deviation)			
			NYSE	NASDAQ	Other	
<i>Actual portfolio</i>						
What percentages (approximately) of the companies that are in your portfolio are trading on the (1. New York Stock Exchange (NYSE); 2. NASDAQ; 3. Other Exchanges	1	35	69.0(28.31)	23.00(21.67)	8.00(21.39)	
	2	33	69.09(24.86)	22.42(17.55)	8.64(23.92)	
	T		-0.01	0.12	-0.12	
	Pr > t		0.9888	0.9049	0.9082	
		W ^a	0.7098	0.8521	0.5353	
What percentage (approximately) of the companies that are in your portfolio in terms of dollar value, is trading on? (1. New York Stock Exchange (NYSE); 2. NASDAQ; 3. Other Exchanges	1	35	65.00(30.17)	23.14(22.20)	11.86(25.30)	
	2	33	71.82(23.88)	19.70(15.41)	8.48(23.93)	
	T		-1.03	0.74	0.56	
	Pr > t		0.3071	0.4623	0.5747	
		W	0.3280	0.6455	0.8459	
	Group	N	Yes	No	Pearson χ^2	p-value
<i>Industry</i>						
Is there any industry that you tend to avoid investing in? (Yes/No)	1	35	46%	54%	0.6131	0.4336
	2	33	36%	64%		

Notes: Group 1 – Disclosure treatment (i.e., pro forma net income disclosed in notes to financial statements). Group 2 – Recognition treatment (i.e., recording in the income statement of the expense related to stock options).

^aWilcoxon 2 sample test: two-sides normal approximation.

Table 2. Statistics on Manipulation Questions Cross-Tabulation of Format and Acknowledgement Regarding Pro Forma Disclosure, Income Statement Expense for Stock Options and Net Income Trend.

	Group	N	Yes	No	Pearson χ^2	p-value	
<i>Pro forma disclosure</i>							
Did ABC report a pro forma figure in its 2003 note on "Stock options" (Yes/No)	1	35	100%	0%	3.3287	0.0681	
	2	33	9%	91%			0.1089 ^a
<i>Expense relative to stock options</i>							
Did ABC record a "Stock options" expense in its 2003 financial statements? (Yes/No)	1	35	20%	80%	25.9636	<0.0001	
	2	33	82%	18%			
	Group	N	<	=	>	Pearson χ^2	p-value
<i>Net income trend</i>							
ABC's reported net income for fiscal year 2003 that was _____ than its reported net income in fiscal year 2002 (inferior, stable, superior)	1	35	6%	0	94%	46.2270	<0.0001
	2	33	88%	0	12%		

Notes: Group 1 – Disclosure treatment (i.e., pro forma net income disclosed in notes to financial statements). Group 2 – Recognition treatment (i.e., recording in the income statement of the expense related to stock options).

^aFisher Exact test two-sided.

reported net income was superior (inferior) than its 2002 reported net income. Overall, the results of these questions indicate that the manipulation was perceived as intended.

Tests on Underlying Judgments (H_1)

The results for questions pertaining to the underlying judgments are presented in [Table 3](#). As mentioned before, the underlying judgments were measured using questions relating to the perceived risk (Panel A), expected accounting return (Panel B), and the overall perception (Panel C). As Panel A of [Table 3](#) reveals the responses to the questions relating to the perceived risk, which consist of the (1) perceived investment risk, (2) overall financial condition, and (3) ability to pay its debts. The perceived investment risk is more pessimistic (greater) for the recognition treatment (mean 7.33) than for the disclosure treatment (mean 6.49). The perceived overall financial condition is also more pessimistic (lower) for the recognition treatment (mean 4.61) than for disclosure treatment (mean 5.43). Similarly, the judgments on the company's ability to pay its debts are more pessimistic (lower) for the recognition treatment (mean 4.76) as opposed to the disclosure treatment (mean 5.71). In all three questions, the differences between the disclosure treatment and the recognition treatment are highly significant (with respective p -values of 0.002, 0.012, and 0.005).

Panel B of [Table 3](#) reveals the responses obtained for the questions pertaining to expected accounting return which consists of the (1) expected return on equity, (2) future net income direction, and (3) growth prospects. The results for expected return on equity show that in all three yearly forecasts expected return on equity were smaller for the recognition treatment relative to the disclosure treatment. The difference is statistically significant at the 1% level for each year. The response to future net income direction shows that 34% predicted decrease in net income for the disclosure treatment as opposed to 31% indicating increase and 34% indicating no change. For the recognition treatment 55% predicted decrease as opposed to 12% predicting increase and 33% indicating no change. A χ^2 test indicated that the differences were statistically significant between the groups (p -value 0.038). The responses to growth prospects show that the company's anticipated growth prospects are more pessimistic (smaller) for the recognition treatment (mean 4.30) relative to the disclosure treatment (mean 4.80). The results from a Wilcoxon test show that the difference is marginally significant between the groups (p -value 0.058).

Table 3. Tests on Underlying Judgments of the Company (H_1).

Panel A: Perceived risk						
	Group	<i>N</i>	Mean	Standard deviation	<i>t</i> -test (<i>p</i> -value)	Wilcoxon (<i>p</i> -value)
<i>1. Perceived investment risk</i>						
I believe that an investment in ABC Inc. is: ____ (on a 9 points Likert scale where 1 represents very low risk and 9 represents very high risk)	1	35	6.49	1.40	0.005	0.002
	2	33	7.33	1.22		
<i>2. Overall financial condition</i>						
I believe ABC's overall financial condition is: _____(on a 9 points Likert scale where 1 represents a very poor and 9 represents very good)	1	35	5.43	1.63	0.015	0.012
	2	33	4.61	1.39		
<i>3. Ability to pay its debts</i>						
I believe ABC's ability to pay its debts as they come due is: _____(on a 9 points Likert scale where 1 represents a very poor and 9 represents very good)	1	35	5.71	1.62	0.008	0.005
	2	33	4.76	1.54		
Panel B: Expected accounting return						
	Year	Mean (Standard deviation)		<i>t</i> -test (<i>p</i> -value)	Wilcoxon (<i>p</i> -value)	
		Group 1	Group 2			
<i>1. Expected return on equity</i>						
What accounting return on equity do you expect for the next 3 years? ___%	2004	7.86	3.58	<0.001	<0.001	
	2005	8.17	5.18	0.002	0.001	
	2006	8.34	5.85	0.014	0.013	

	Group	N	↓	No change	↑	$\chi^2(p\text{-value})$
<i>2. Future net income direction</i>						
I believe ABC's future Net Income will:	1	35	34%	34%	31% ^a	0.038
____ (decrease, no change, increase)	2	33	55%	33%	12%	
	Group	N	Mean	Standard deviation	t-test (p-value)	Wilcoxon (p-value)
<i>3. Growth prospects</i>						
I believe ABC's growth prospects are:	1	35	4.80	1.66	0.102	0.058
____ (on a 9 points Likert scale where 1 represents very unfavorable and 9 represents very favorable)	2	33	4.30	1.53		
Panel C: Overall perception						
	Group	N	Mean	Standard deviation	t- test (p-value)	Wilcoxon (p-value)
<i>1. Perceived investment attractiveness</i>						
I believe that an investment in ABC shares is: _____ (on a 9 points Likert scale where 1 represents not at all attractive and 9 represents very attractive)	1	35	4.46	2.01	0.002	0.002
	2	33	3.15	1.50		

^aThe row does not total 100% due to rounding.

Panel C of Table 3 reveals the responses obtained to the overall perception of the attractiveness of this investment. On average, results showed that participants of the recognition treatment (mean 3.15) felt an investment in ABC shares was less “attractive” relative to the disclosure treatment (mean 4.46). The difference was statistically significant (p -value 0.002).

In summary, the results of Table 3 show that, in the presence of recognition versus disclosure, participants estimate more pessimistic underlying judgments (higher perceived risk, lower expected accounting return, less favorable overall perception) about the firm when it recognizes the cost of stock options as an expense. These results all provide evidence to support H_1 which postulates that an individual’s overall underlying judgment of the company will be more pessimistic when stock-option costs are recognized in the income statement than when they are disclosed in a footnote.

Tests on Primary Judgments (H_2 and H_3)

As mentioned before, the primary judgments pertain to the (1) expected future stock price direction and (2) stock price value. The results are shown in Table 4. The mean expected future direction in stock price for participants in the recognition treatment (3.15) was more pessimistic (lower) than the mean for participants in the disclosure treatment (4.80). The difference was statistically significant (p -value < 0.001). This provides evidence to support H_2 which postulates that investors will predict a more pessimistic expected future stock price direction when stock-option costs are recognized in the income statement than when they are disclosed in a footnote.

Similarly, the mean estimated stock price value of participants in the recognition treatment (\$7.30) was lower than the estimated mean for participants in the disclosure treatment (\$24.43). The difference was statistically significant (p -value 0.001). This provides evidence to support H_3 which postulates that investors will estimate a lower stock price value when stock-option costs are recognized in the income statement then when they are disclosed in a footnote.

Supplementary Tests on the Perceived Quality of Accounting Information

We continued our analysis in order to investigate if users’ perceived quality of accounting information is influenced by the stock-option reporting format (recognition versus disclosure). This perceived quality was measured

Table 4. Tests on Primary Judgments (H₂ and H₃).

	Group	N	Mean	Standard deviation	t-test (p-value)	Wilcoxon (p-value)
<i>Expected future stock price direction</i>						
Assume that the financial information for 2003 has just been publicly released and that the stock market is receiving this information for the first time. Please indicate what you believe will most likely happen to the stock price of ABC upon release of this information: _____ (on a 9 points Likert scale where 1 represents significant price decrease and 9 represents significant price increase)	1	35	4.80	1.86	< 0.001	< 0.001
	2	33	3.15	1.30		
<i>Stock price value</i>						
Assume that ABC's financial information for 2003 has just been publicly released. Please indicate what you believe to be a fair price for one share of ABC stock: _____\$	1	35	\$24.43	22.86	0.001	0.001
	2	33	\$7.30	8.66		

Notes: Group 1 – Disclosure treatment (i.e., pro forma net income disclosed in notes to financial statements). Group 2 – Recognition treatment (i.e., recording in the income statement of the expense related to stock options).

using five questions relating to the financial statement usefulness, limpidity, reliability, clarity, and confidence in valuation judgment. As with previous measures, a nine-point Likert scale is used for each question. The results for this final section are presented in Table 5. Only one perception, financial statement usefulness, appears to be influenced by the reporting format. Financial statement usefulness is perceived marginally higher for the recognition treatment (5.88) than for the disclosure treatment (5.03) with

Table 5. Tests on the Perceived Quality of Accounting Information.

	<i>G</i>	<i>N</i>	Mean	Standard deviation	<i>t</i> -test (<i>p</i> -value)	Wilcoxon (<i>p</i> -value)
<i>Financial statements usefulness</i>						
In determining the company's true, overall financial performance, I believe ABC's financial statements are: _____ (on a 9 points Likert scale where 1 represents not at all helpful and 9 represents very helpful)	1	35	5.03	1.92	0.029	0.077
	2	33	5.88	1.69		
<i>Limpidity of financial statements</i>						
I believe ABC's financial statements are: _____ (on a 9 points Likert scale where 1 represents not at all confusing and 9 represents very confusing)	1	35	4.46	2.08	0.450	0.465
	2	33	4.52	1.70		
<i>Reliability of financial statements</i>						
I believe ABC's financial statements are: _____ (on a 9 points Likert scale where 1 represents very unreliable and 9 represents very reliable)	1	35	5.66	1.85	0.329	0.182
	2	33	5.48	1.28		
<i>Clarity of financial statements</i>						
I believe ABC's financial statements are: _____ (on a 9 points Likert scale where 1 represents not at all clear and 9 represents very clear)	1	35	5.51	2.08	0.473	0.351
	2	33	5.48	1.48		
<i>Confidence in valuation judgment</i>						
Please assess how confident you are in your valuation judgment: _____ (on a 9 points Likert scale where 1 represents not confident and 9 represents very confident)	1	35	5.14	2.38	0.313	0.258
	2	33	4.88	2.04		

Notes: Group 1 – Disclosure treatment (i.e., pro forma net income disclosed in notes to financial statements). Group 2 – Recognition treatment (i.e., recording in the income statement of the expense related to stock options).

a p -value of 0.029. This result indicates that the financial statements were seen more helpful in the presence of stock-option costs recognition as opposed to a note disclosure. None of the other qualities (limpidity, reliability, clarity, and confidence in valuation judgment) was statistically different between the two experimental groups.

In summary the results indicate that financial statement users make more pessimistic underlying judgments (higher perceived risk, lower expected accounting return, worse overall perception) and primary judgments (negative future stock price direction and lower stock price valuation) when the cost associated with stock options is recognized. The results also show that financial statements are marginally more helpful in the presence of recognition versus disclosure. We also note, however, that when questioned on the financial statements overall, the participants did not find recognition of stock options to make the financial statements significantly less confusing, or to significantly increase the reliability and clarity of financial statements. Simply recognizing stock-option costs may not have been sufficient to significantly change their viewpoints on the qualities of financial statements even if it influenced significantly their underlying and primary investment judgments.

CONCLUSIONS

Most of the prior research has examined the consequences of disclosing stock-option costs by means of stock market studies of the impact of pronouncements on share prices. Recent experimental research (Belzile et al., 2006; Viger et al., 2008) examined the judgments of financial statement users when stock options are recognized versus disclosed in the notes. These studies conclude that both unsophisticated investors and professional loan officers are influenced by the reporting format. Our results indicate that sophisticated financial statements users are susceptible to the format effect in making their personal investment judgments. Thus, our findings have broad policy implications as they corroborate the stance of the FASB that disclosure is not a substitute to recognition.

As a regulatory body, FASB must be concerned with the informational content for individuals, not just the whole market. In point of fact, Donald T. Nicolaisen (who was appointed as the Chief Accountant of the SEC in August 2003) commented that investors deserve financial statements that are more transparent and easier to read. While talking in general terms (though his statement is applicable to stock-option reporting as well), he stated that

in some cases excessively detailed accounting rules have given those who choose to abuse the system a means to achieve better looking numbers in the financial statements by circumventing the fundamental principles behind the rules. Not recognizing stock-option expenses as was allowed prior to the introduction of SFAS 123R is a perfect example of this form of circumventing. He noted that a dominant item on his agenda was helping *all* investors (by implication large and small) to be better informed. Subsequently, a year later he noted that it was important to know if the “low hanging fruit” had in fact been plucked before proceeding to other issues which need to be addressed (refer to Rouse, 2004 for more details on statements made by Nicolaisen). Our results suggest that the SFAS 123R mandatory recognition may prevent undesired consequences due to faulty acquisition or/and processing of information on stock-option compensation.

This study has some limitations. First, our final sample was drawn by means of two different sampling procedures. It can be criticized as not representing the population of IMA members at large. Second, the experimental design did not consider all the costs and benefits associated with investors’ judgments. Participants, for example, were provided with less information than is normally available. The data provided was mainly financial information and could be analyzed in approximately half an hour. In real life, investors may take more than thirty minutes, may use additional non-financial information, and likewise, may consult with colleagues in arriving at their judgments. Third, this study did not analyze the investment decision resulting from the many ambiguous answers received to this question. Because this study is limited to individual judgments, it would be appropriate to evaluate the economic impact of SFAS 123R by examining the behavior of stock prices. Results from such research would allow the FASB to corroborate the impact of the mandatory recording of stock options.

NOTES

1. The SFAS 123R is in accordance with the international standard on stock-based payments (International Accounting Standards Board (IASB), 2004), which was adopted for application in the financial statements of fiscal periods beginning on or after January 1, 2005. The International Accounting Standard requires recognition of stock-based compensation as an expense at grant date by using a fair value method.

2. The option valuation models in use today were not yet developed.

3. If the exercise price is equal to or greater than the market price at the granting date, then the intrinsic value of stock options is equal to zero.

4. Question one also requested whether participants would be willing to invest in the company however, quite a number answered this question ambiguously using the words “perhaps” or “it depends.” Due to the many ambiguous or equivocal answers, we could not include responses to this question in our statistical tests.

ACKNOWLEDGMENTS

This paper was presented and accepted as part of the 3rd International Conference on Enterprise Systems and Accounting (ICESAcc'06) 26–27 June 2006, Santorini Island, Greece. The authors gratefully acknowledge the editor who gave insightful and constructive comments.

REFERENCES

- Aboody, D., Barth, M. E., & Kasznik, R. (2004). SFAS no. 123 stock based compensation expense and equity market values. *The Accounting Review*, 79(2), 251–275.
- American Accounting Association Financial Accounting Standards Committee. (2005). Response to the FASB's exposure draft on share-based payment: An amendment of FASB statements no. 123 and no. 95. *Accounting Horizons*, 19(2), 101–114.
- Anandarajan, A., Viger, C., & Curatola, A. P. (2002). The information content of alternative going concern reports: A Canadian experience. *Canadian Accounting Perspectives*, 1(2), 141–162.
- Bailey, W. T. (1981). The effects of audit reports on chartered financial analysts' perceptions of the sources of financial statement and audit report messages. *The Accounting Review*, 55, 882–896.
- Balsam, S., Bartov, E., & Yin, J. (2004). *Disclosure versus recognition of option expense: An experimental investigation of SFAS no. 148 and stock returns*. Unpublished working paper. Temple University.
- Bamber, E. M., & Stratton, R. A. (1997). The information content of the uncertainty-modified audit report: Evidence from bank loan officers. *Accounting Horizons*, 11(2), 1–11.
- Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In: G. Constantinides, M. Harris & R. Stulz (Eds), *Handbook of the economics of finance*. Amsterdam, The Netherlands: Elsevier/North-Holland.
- Barth, M. E. (1991). Relative measurement errors among alternative pension asset and liability measures. *The Accounting Review*, 66(3), 433–464.
- Bell, T., Landsman, W., Miller, B., & Yeh, S. (2002). The valuation implication of employee stock option accounting for profitable software firms. *The Accounting Review*, 77(4), 971–996.
- Belzile, R., Fortin, A., & Viger, C. (2006). Recognition versus disclosure of stock option compensation: An analysis of judgments and decisions of non-professional investors. *Canadian Accounting Perspectives*, 5(2), 147–180.

- Bernard, V., & Schipper, K. (1994). *Recognition and disclosure in financial reporting*. Unpublished working paper. University of Michigan and University of Chicago.
- Bloomfield, R. J. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons*, 16(3), 233–243.
- Choi, B., Collins, D. W., & Johnson, W. B. (1997). Valuation implications of reliability differences: The case of nonpension postretirement obligations. *The Accounting Review*, 72(3), 351–383.
- Elias, R. Z., & Johnston, J. G. (2001). Is there incremental information content in the going concern explanatory paragraph? *Advances in Accounting*, 18, 105–117.
- Fiske, S., & Taylor, S. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Frederickson, J. R., & Miller, J. S. (2004). The effects of proforma earnings disclosures on analysts' and nonprofessional investors' equity valuation judgments. *The Accounting Review*, 79(3), 667–686.
- Gul, F. A. (1987). The effects of uncertainty reporting on lending officers' perception of risk and additional information required. *ABACUS*, 23(2), 172–179.
- Herz, R. (2004). Testimony before the subcommittee on financial management, the budget and international security. U.S. Congress. Available at <http://www.fasb.org/testimony/04.20.04>.
- Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure and financial reporting. *Journal of Accounting and Economics*, 36(December), 337–386.
- Hirst, D. E., & Hopkins, P. E. (1998). Comprehensive income reporting and analysts' valuation judgments. *Journal of Accounting Research*, 36(Suppl.), 47–75.
- Hirst, E., Hopkins, P., & Wahlen, J. (2004). Fair values, income measurement, and bank analysts' risk and valuation judgments. *The Accounting Review*, 79(2), 453–472.
- Ijiri, Y., Jaedicke, R. K., & Knight, K. E. (1966). The effects of accounting alternatives on management decisions. In: R. K. Jaedicke, Y. Ijiri & O. Nielson (Eds), *Research in accounting measurement* (pp. 186–199). Sarasota, FL: American Accounting Association.
- International Accounting Standards Board (IASB). (2004). *International financial reporting standards (IFRS). Share-based payment*. London: IASB.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Lasalle, R. E., & Anandarajan, A. (1997). Bank loan officers' reactions to audit reports issued to entities with litigation and going concern uncertainties. *Accounting Horizons*, 11(2), 33–40.
- Libby, R. (1979). The impact of uncertainty reporting on the loan decision. *Journal of Accounting Research*, 17(Suppl.), 35–57.
- Libby, R., Bloomfield, R., & Nelson, M. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society*, 27(8), 775–810.
- Maines, L. A. (1995). Judgment and decision making research in financial accounting: A review and analysis. In: R. H. Ashton & A. H. Ashton (Eds), *Judgment and decision making research in accounting and auditing*. Cambridge, MA: Cambridge University Press.
- Maines, L. A., & McDaniel, L. S. (2000). Effects of comprehensive-income characteristics on nonprofessional investors' judgments: the role of financial-statement presentation format. *The Accounting Review*, 75(2), 179–207.

- Rouse, R. W. (2004). Stock option controversy continues. *Journal of Corporate Accounting and Finance*, 16(2), 79–81.
- Tinic, S. M. (1990). A perspective on the stock market's fixation on accounting numbers. *The Accounting Review*, 65(4), 781–796.
- Viger, C., Anandarajan, A., Curatola, A. P., & Amar, W. B. (2004). Behavioral implications of alternative going concern reporting formats. *Advances in Accounting Behavioral Research*, 7, 51–71.
- Viger, C., Belzile, R., & Anandarajan, A. (2008). Disclosure versus recognition of stock option compensation: The effect on credit evaluation of loan officers. *Behavioral Research in Accounting*, 20(1), 91–113.

AUDITOR PERFORMANCE VARIATION: IMPACT OF SUB-SPECIALTY KNOWLEDGE DIFFERENCES BETWEEN INDUSTRY-SPECIALISTS

Carlin Dowling and Robyn Moroney

ABSTRACT

The extant literature has established that industry-specialist auditors gain performance-enhancing industry-specific sub-specialty knowledge (e.g., Solomon, Shields, & Whittington, 1999) via training and on the job experience. This knowledge has been shown to allow specialists to outperform non-specialists on a range of industry-specific tasks. The current study extends this line of research by comparing and contrasting the relative performance gains enjoyed by industry-specialist auditors in two different industry settings, one regulated and the other unregulated. When specializing in regulated industries, auditors gain very detailed industry-specific knowledge which is not the case for specialists in unregulated industries (Dunn & Mayhew, 2004). By comparing industry-specialists to non-specialists with matching industry-based experience, this study measures the relative benefits of specialization in different industry settings, rather than the benefits of specialization per se, which has been well established in the literature. This study finds that the

Advances in Accounting Behavioral Research, Volume 11, 59–79

Copyright © 2008 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11003-1

performance gains made by regulated industry-specialists significantly outweigh those made by unregulated industry-specialists on industry-specific tasks. The implications of these results for future research and practice are explored in the body of the chapter.

INTRODUCTION

This chapter explores the relative performance gains available to auditors who specialize in different industry settings. Industry-specialist auditors gain sub-specialty knowledge via training and experience in one industry setting (Bonner & Lewis, 1990; Libby, 1995). This specialist knowledge sets them apart from other auditors and allows them to outperform non-specialist auditors on industry-specific tasks (e.g., Solomon, Shields, & Whittington, 1999). Existing behavioral research has concentrated on the performance gains made by industry-specialists compared to non-specialists. This is an all or nothing comparison; industry-specialists have the relevant performance-enhancing sub-specialty knowledge and non-specialists do not. As a result, industry-specialists across a variety of industries outperform non-specialists who do not have industry experience (Solomon et al., 1999; Taylor, 2000; Owghoso, Messier, & Lynch, 2002) on industry-specific tasks.

This study extends this line of research by comparing industry-specialists to non-specialists with matched industry-based experience in two very different industry settings, one regulated and the other unregulated. As the extent of regulation in an industry increases, more industry-specific sub-specialty knowledge is required to conduct the audit (Dunn & Mayhew, 2004). Consequently, industry-specialists in regulated industries are expected to clearly outperform non-specialists with matched industry-based experience on an industry-specific task. These performance gains are not expected for specialists in an unregulated industry because the extent of industry-specific sub-specialty knowledge is not as great.

This study experimentally examines and compares the task performance of industry-specialists with that of non-specialist auditors with matched industry-based experience. The industry-specialist participants were sourced from first-tier audit firms. The non-specialist participants with matched industry-based experience were sourced from mid-tier audit firms. This choice was imposed by the way these firms structure and utilize their auditors. The audit divisions of the first-tier firms that provided participants are structured along industry lines with most auditors specializing in one

industry. The first-tier auditors who participated in this study are firm-designated industry-specialists in either a regulated or unregulated industry.¹ The audit divisions of the participating mid-tier firms are not structured along industry lines. Auditors in this tier are not firm-designated industry-specialists. Using non-specialist auditors with matched industry-based experience as the performance base enables the assessment of the relative benefits of specialization in different industry settings and provides important insights into when industry-specialists are likely to outperform non-specialists with matched experience.

The findings support the hypotheses and suggest that when compared to non-specialist auditors with matched industry-based experience, significant task performance gains are only made by industry-specialists in a regulated industry setting and not an unregulated industry setting. This finding does not conflict with the current literature that finds that industry specialization enhances auditor performance. The contribution of this study is that it provides evidence that the benefits of industry specialization are not uniform across different industry settings. The findings suggest that in some industry settings (unregulated) the relative benefits of specialization are not as marked as in other industry settings (regulated). The results have implications for studies that use industry specialization as a proxy for audit quality (for a review see Francis, 2004). While industry specialization is a valid proxy for audit quality in regulated industries, the results indicate that, depending on the research question, industry-based experience may be a sufficient proxy for audit quality in unregulated industries.

The remainder of this chapter is divided into four sections. The hypotheses are developed in the next section, followed by the research design. The results are then analyzed, followed by the conclusions and limitations of this study.

BACKGROUND AND HYPOTHESIS DEVELOPMENT

Industry specialization is a form of expertise, which is the ability, acquired by practice, to perform qualitatively well in a particular domain (Bédard & Chi, 1993). An expert is an individual who has acquired domain-specific knowledge through training and experience (Spence & Brucks, 1997) that enables the individual to achieve superior task-specific performance. Industry-specialist auditors outperform non-specialists because they develop “industry-specific skills and expertise over and above normal auditor expertise” (Craswell, Francis, & Taylor, 1995, p. 301).

Substantial research supports the view that industry-specialist audit firms provide better quality audits (Kwon, 1996; Hogan & Jeter, 1999; DeFond, Francis, & Wong, 2000; Balsam, Krishnan, & Yang, 2003; Krishnan, 2003; Francis, 2004; Francis, Reichelt, & Wang, 2005). While audit firm size and reputation are important determinants of audit quality, industry specialization has been used as a way to differentiate between first-tier firms (Kwon, 1996; Francis, Stokes, & Anderson, 1999). The use of an industry-specialist audit firm is associated with higher earnings response coefficients, higher quality voluntary disclosures, lower discretionary accruals, lower levels of fraudulent reporting and higher market valued earnings surprises (Balsam et al., 2003; Krishnan, 2003; Dunn & Mayhew, 2004; Francis et al., 2005). While these studies have focused on understanding industry specialization at the audit firm level, behavioral studies complement this stream of research by investigating industry specialization at the auditor level.

According to behavioral decision theory, performance is determined by an individual's experience, ability and knowledge (Bonner & Lewis, 1990; Libby & Luft, 1993; Libby & Tan, 1994; Libby, 1995). Expertise, which combines knowledge and experience (Bonner & Lewis, 1990), has been defined as the ability, acquired by practice, to perform qualitatively well in a particular domain (Bédard & Chi, 1993). Expertise is more than gaining general experience, it is the knowledge acquired directly through relevant experiences and indirectly through education and training (Bédard, 1989).

Bonner and Lewis (1990) criticized early behavioral studies that predicted superior auditor performance based upon years of experience or rank alone because these studies were confusing experience with expertise. They argued that years of experience and rank do not capture the nature of the specific experiences of individual auditors. Although years or rank capture an individual's experience working as an auditor, they do not measure the knowledge an auditor amasses from their experiences within a specific domain. Bonner and Lewis (1990) differentiated between the different types of knowledge that can be gained by auditors. General domain knowledge is the knowledge gained by most auditors via instruction and on the job experience. It is the knowledge common to most, if not all, auditors. General domain knowledge is important for all auditors.

Another type of knowledge is sub-specialty knowledge (Bonner, 1990; Bonner & Lewis, 1990; Bedard & Biggs, 1991b; Bédard & Chi, 1993; Libby & Tan, 1994; Wright & Wright, 1997). Like general domain knowledge, sub-specialty knowledge is gained via instruction and on the job experience, but unlike general domain knowledge, sub-specialty

knowledge is only gained by people who work within a particular sub-specialty area, such as auditing clients within one industry setting. It is the accumulation of domain-specific sub-specialty knowledge which improves performance on tasks set within that domain. As a result, industry-specialists are found to perform better than non-specialists on tasks set within their industry. This does not mean that industry-specialists are better auditors or have superior general domain knowledge; their experiences auditing clients predominantly within one industry setting allows them to perform better on tasks set within that industry setting (Solomon et al., 1999).

There are different kinds of sub-specialty knowledge. For example, Bedard and Biggs (1991a) found that some, but not all, first-tier managers outperformed seniors when explaining a recognized pattern of cues during an analytical procedures task. This superior performance was attributed to task-based experience (i.e., experience in recognizing financial statement errors) rather than manufacturing industry-based experience, as all managers had the relevant industry experience. In this case, the sub-specialty knowledge was improved by task-specific rather than industry-specific experience.

Bedard and Biggs (1991b) and Wright and Wright (1997) found that manufacturing and retail industry experience, respectively, improve performance on a hypothesis generation task but not necessarily on other tasks (Wright & Wright, 1997). These studies focus on error frequency knowledge during a hypothesis generation task. The current study uses auditing tasks that have been developed to test industry-specific knowledge. They are multifaceted, which allows us to examine performance over a range of tasks, which is important when testing expertise (Marchant, 1990; Simnett, Lockett, & Wright, 2000).

Biggs, Selfridge, and Krupka, (1993) stress the importance of task-based experience in the development of expertise. They found that task performance in turn is enhanced by detailed knowledge of the client, including the industry in which it operates. As such, industry experience provides important background knowledge for clients that operate in that industry, which in turn enhances task performance.

A number of recent behavioral studies have reported superior performance on a range of tasks by first-tier industry-specialist auditors. These studies have found that industry-specialists have more accurate non-error frequency knowledge (Solomon et al., 1999), are more effective when assessing inherent risk levels for accounts specific to the industry in which they specialize (Taylor, 2000), are more effective when detecting errors for clients in their industry (Owhoso et al., 2002), develop task knowledge

differently depending on their area of specialization (Thibodeau, 2003), are able to assess audit risk more accurately (Low, 2004) and are able to interpret and complete partial cue patterns (Hammersley, 2006). These studies compared first-tier auditors working in and out of specialization.

Although a few studies investigating industry specialization at the audit firm level have recognized that industries are not homogenous (see, e.g., Dunn & Mayhew, 2004; Godfrey & Hamilton, 2005), behavioral studies have assumed that industries are homogenous, and that the benefits of specialization accrue equally across all industries. The problem with this assumption is that industries are not homogenous. One important difference is the extent to which an industry is subject to regulation (Abdolmohammadi, Seafoss, & Shanteau, 2004).

This study extends behavioral research investigating industry specialization by comparing the relative benefits of industry specialization in a regulated versus an unregulated industry setting. Differences in the extent of regulation influence the level of sub-specialty knowledge required to audit clients in specific industries. This study argues that the extent of regulation results in different performance gains available to auditors who specialize in a regulated industry relative to those available to auditors who specialize in an unregulated industry.

Less domain-specific sub-specialty knowledge is required when auditing clients in unregulated industries such as retail, service and manufacturing than when auditing clients in regulated industries (Dunn & Mayhew, 2004; Abdolmohammadi et al., 2004). The lower level of regulation means that non-specialists with unregulated industry experience have the opportunity to acquire the industry-specific knowledge required to audit clients in that industry, which will enable them to perform at a comparable level with auditors who specialize in that industry. This discussion leads to the following hypothesis:

H_{1(Null)}. The performance level of industry-specialist auditors will not be significantly different to non-specialist auditors with matched industry experience for a task set in an unregulated industry.

In contrast, highly regulated industries such as banking, insurance, finance and pension funds require compliance with specific highly detailed reporting requirements. Auditors require detailed sub-specialty knowledge to audit clients in highly regulated industries. The high level of detailed sub-specialty knowledge required to audit clients in highly regulated industries means that auditors who are not specialists in the industry are less likely to acquire the detailed sub-specialty knowledge required to perform at a level

comparable with auditors who specialize in the same regulated industry. Therefore, auditors who are specialists in a regulated industry are likely to obtain significant performance gains over auditors with matched industry experience but who are not industry-specialists. This discussion leads to the following hypothesis:

H₂. The performance level of industry-specialist auditors will be significantly greater than non-specialist auditors with matched industry experience for a task set in a regulated industry.

The first two hypotheses suggest that the task performance gains of industry specialization accrue differentially depending upon the extent of regulation in the industry in which an auditor is a specialist. Specifically, the arguments above imply that the task performance benefits that accrue to industry-specialists in regulated industries exceed the benefits which accrue to specialists in an unregulated industry, when compared with non-specialists who have matched industry experience. This discussion leads to the following hypothesis:

H₃. The difference in task performance levels between industry-specialist and non-specialist auditors with matched industry experience will be positive and significantly greater for a task set in a regulated industry than for a task set in an unregulated industry.

RESEARCH DESIGN AND METHODOLOGY

An experiment was conducted in the offices of all four first-tier and eight mid-tier accounting firms. The industries selected to represent the regulated and unregulated industries were chosen by the participating first-tier firms (see note 1). The pension fund industry was selected as the regulated industry and the manufacturing industry was selected as the unregulated industry. These industries were chosen for their stark differences in complexity and the amount of detailed industry-specific knowledge required when auditing clients in each industry setting.

The participating auditors are from firms that service the same markets and compete for many of the same clients but differ in the way they are structured. Auditors in first-tier firms are allocated to a specific area of industry specialization and they spend most, if not all, of their time working in the one industry and thus have the opportunity to develop highly specialized knowledge and skills (Craswell et al., 1995; Wright & Wright,

1997; Solomon et al., 1999; Taylor, 2000; Owhoso et al., 2002). These auditors are industry-specialists; they receive specific industry training and spend a substantial amount of their time auditing clients in that industry. The first-tier firms provided industry-specialists in the manufacturing and pension fund industries. Consistent with prior research (see, e.g., Solomon et al., 1999; Owhoso et al., 2002), these auditors are firm-designated industry-specialists. In contrast, auditors employed in the participating mid-tier firms audit clients across several industries and have more varied experiences; they are not designated industry-specialists. The non-specialists used in this study were selected based upon their self-reported experience in either the manufacturing or pension fund industries.² Using non-specialists with recent industry experience provides a performance base enabling a comparison of the relative benefits of specialization in different industry settings.

The Experiment

The present study uses the data for a task completed by each participant in either the manufacturing industry or the pension fund industry.³ These tasks were developed with the aid of industry-specialist first-tier audit partners (an expert panel). The tasks are based upon real problems faced by auditors, and were developed to test industry-based knowledge.⁴ Before conducting the experiment, audit partners from all participating firms confirmed that the experimental tasks reflect the type of issues their staff may encounter on audits of clients in the manufacturing and pension fund industries.

The manufacturing task (one page long) deals with research and development expenditure. The task comprises some information regarding a client and five questions. The questions include how to audit research and development expenditure, which items from a list provided in the task material, could be capitalized and how a government grant should be handled. To answer the problem set in the case correctly, participants required knowledge specific to the manufacturing industry, including the impact of technological change on the valuation of inventory and the costing of research and development expenditure.

The pension fund task (two pages long) deals with the identification of audit procedures necessary to confirm listed audit assertions and compliance with legislation. An audit report and a list of investments to be audited are included in the task material. To answer the problem set in the case correctly, participants require knowledge of the regulations impacting the reporting of results in financial statements. The information cues provided

to participants included excerpts from industry-specific accounting standards and regulations. These regulations detail the auditing and disclosure rules particular to pension funds.

An experienced researcher was present when the experiment was conducted. A standard introduction and consent form was provided to each participant outlining in broad terms the purpose of the experiment. Using a unique username and password, participants were then invited to log on to the Internet-based program designed by a specialist programmer. After logging on, each participant read an introduction which explained the project in broad terms and how the software was to be used. Participants then read through a practice task set in the airline industry and accompanying information cues to become familiar with the format of the software. To save time, participants were asked not to provide an answer for the practice task. The participants then completed the manufacturing and pension fund case studies (tasks) discussed previously.

Fig. 1 provides an overview of the experiment. After reading the task materials, participants could either go to the information menu, where a list of information cues was provided, or they could go straight to the solution input screen where they provided their answer.⁵ Participants could move between the task materials, information cues and solution input screen as frequently as they desired before submitting their final solution for the task.

Performance level is measured as the degree of completeness of each participant's solution. The expert panel provided a model solution for the task in their area of expertise. Two expert coders individually blind coded each participant's solution by comparing it to the model solution.⁶ The greater the consistency between a participant's answer to the model solution, the higher a participant's score.

Participants

A total of 86 (48 manufacturing and 38 pension fund) industry-specialist auditors and 83 non-specialist auditors with more than two years audit experience participated in the experiment.⁷ To test the hypotheses, only auditors with experience in the manufacturing or pension fund industry were included in the analysis. Industry-based experience for the non-specialist auditors was measured conservatively as having recently (i.e., during the year leading up to the experiment) audited clients in the manufacturing or pension fund industry. Of the 83 non-specialist auditors, 56 had recent experience in the manufacturing industry and 21 in the pension fund industry. Of these

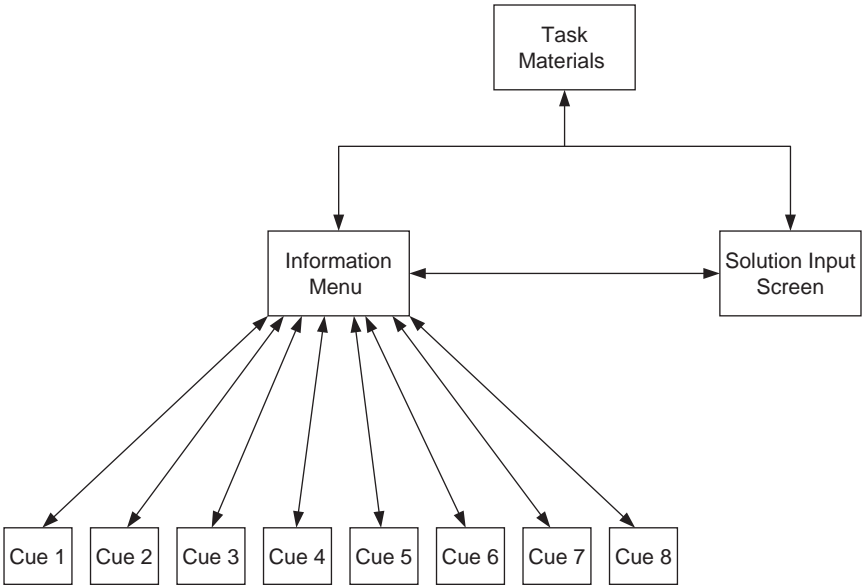


Fig. 1. Overview of the Experiment.

participants, 12 had recent experience in both industries. Testing H_2 requires the four sub-samples of auditors to be independent. Therefore, these 12 auditors were excluded from the manufacturing sub-sample, leaving a total of 44 non-specialist auditors with manufacturing industry experience and 21 with pension fund industry experience.⁸ Both groups of non-specialists reported working similar proportions of the prior year in the relevant industry (27% for the manufacturing and 25% for the pension fund non-specialists ($t = .385, p > .7$, not tabulated)).

Table 1 contains the descriptive statistics for the participants. On average the non-specialist auditors had more audit experience than the industry-specialist auditors. Table 1, Panel A shows that manufacturing industry-specialists had an average of 5 years experience compared with 6.7 years for the non-specialist auditors with manufacturing industry experience ($t = 2.583, p < .02$, not tabulated). Table 1, Panel B shows that pension fund industry-specialists had an average of 7.5 years experience compared with 8 years for the non-specialist auditors with pension fund industry experience ($t = .280, p > .7$, not tabulated). Table 1 also shows that similar proportions of auditors reported having experience with the experimental task; 46% of industry-specialists and 54% of non-specialists working on the

Table 1. Descriptive Statistics for the Participants.

<i>Panel A: Manufacturing (H₁)</i>		
	Specialists in the Manufacturing Industry	Non-specialists with Manufacturing Industry Experience
Number of participants	48	44
Average number of years of audit experience	5 yrs	6.7 yrs
Minimum/Maximum number of years of audit experience	Minimum 2 yrs/ Maximum 16 yrs	Minimum 2 yrs/ Maximum 18 yrs
Proportion of participants with personal experience advising clients on issue used in the manufacturing task (research and development)	46%	54%
<i>Panel B: Pension Fund (H₂)</i>		
	Specialists in the Pension Fund Industry	Non-Specialists with Pension Fund Industry Experience
Number of participants	38	21
Average number of years audit experience	7.5 yrs	8 yrs
Minimum/Maximum number of years of audit experience	Minimum 2 yrs/ Maximum 27 yrs	Minimum 2 yrs/ Maximum 30 yrs
Proportion of participants with personal experience advising clients on issue used in the pension fund task (investments for a pension fund)	97%	95%

manufacturing task ($t = .717$, $p > .4$, not tabulated) and 97% of industry-specialists and 95% of non-specialists working on the pension fund task ($t = .426$, $p > .6$, not tabulated) had task experience. Thus task experience levels were consistent between the groups of auditors being compared.⁹

All industry-specialist participants work in first-tier firms and all non-specialist participants work in mid-tier firms. This choice was made as most first-tier auditors are industry-specialists and mid-tier auditors are not

specialists.¹⁰ The audit divisions in first-tier firms are set up along industry lines with most staff being firm-designated industry-specialists, receiving training and audit experience within one industry setting. The audit divisions of the mid-tier are not set up in the same way. Staff at those firms receive training and audit experience in a variety of industry settings. This distinction is an important one for this study, as the aim is to uncover when this investment in industry-specific training and experience is most advantageous.

RESULTS

For reasons previously stated, all industry-specialist participants were obtained from first-tier firms and all non-specialist participants were obtained from mid-tier firms.¹¹ Before testing the results of the hypotheses it is important to establish whether firm size interferes with the interpretation of the results. An ANOVA was run, comparing auditor performance of all participants from the first-tier and all participants from the mid-tier. The results are not significant ($F = .421, p > .5$, not tabulated). Thus, firm size does not appear to be driving the results reported below.¹²

The results presented in Tables 2 and 3 are consistent with H_1 and H_2 . H_1 predicted that industry-specialists will not significantly outperform non-specialists with matched industry experience in an unregulated (manufacturing) industry. Table 2 shows that manufacturing industry-specialists (mean 4.94) outperform non-specialists (mean 4.59) with matched industry experience but as hypothesized the difference is not significant ($F = 2.232, p > .1$, two tail) after controlling for audit ($F = 3.368, p > .07$) and task experience ($F = 1.108, p > .2$). The results are consistent with the expectation (H_1) that industry-specialist auditors do not significantly outperform their non-specialist counterparts on a task set in an unregulated industry.¹³

H_2 predicted that industry-specialists will significantly outperform non-specialists with matched industry experience in a regulated (pension fund) industry. Table 3 shows that industry-specialists (mean 3.55) significantly outperform non-specialist auditors (mean 2.62) with recent pension fund industry experience ($F = 5.215, p < .03$, two tail) after controlling for audit ($F = .068, p > .7$) and task experience ($F = .289, p > .5$). These findings support H_2 . In a regulated industry that requires significant sub-specialty knowledge, the task performance of industry-specialists is significantly higher than their non-specialist counterparts.

H_3 predicted that the difference in task performance levels between industry-specialist and non-specialist auditors with matched industry

Table 2. Manufacturing Industry.

<i>Panel A: ANOVA for Manufacturing Industry Performance Level (H_1)</i>				
		<i>F</i>	<i>p</i> ^a	
Specialists versus non-specialists with industry experience		2.232	.139	
Covariates				
Audit experience		3.368	.070	
Industry experience		1.108	.295	
<i>Panel B: Mean (Standard Deviation) [Range]</i>				
Specialists (<i>n</i> = 48)	Non-Specialists with Industry Experience (<i>n</i> = 44)	<i>t</i> ^b	<i>p</i> ^c	
4.94 (1.88) [1.0–9.5]	4.59 (1.76) [0.0–8.0]	.911	.365	

^aTwo tail.

^bAs Levene's test for equality of variances is not significant ($F = .012, p > .9$) equal variances are assumed. Shapiro–Wilk test of normality (for sample sizes less than 50) indicates that the industry-specialist ($SW = .979, p > .5$) and non-specialist ($SW = .975, p > .4$) groups are normally distributed.

^cOne tail.

experience will be positive and significantly greater for a task set in a regulated industry compared with a task set in an unregulated industry. Fig. 2 presents a visual diagram of this comparison using the means reported in Tables 2 and 3.¹⁴ From Fig. 2, it can be seen that the slope of the line between the mean performance level for sub-samples 1 and 2 who undertook the manufacturing task is less steep than the slope of the line between the mean performance level for sub-samples 3 and 4. In contrast, the slope of the difference in the mean performance level for sub-samples 3 and 4, who undertook the pension fund task, is steep, indicating a significant performance difference. The difference in the steepness of the two slopes depicted in Fig. 2 is consistent with H_3 .

To test H_3 , an ANOVA was run on the four sub-samples depicted in Fig. 2.¹⁵ A significant interaction is found between specialization and task ($F = 8.122, p < .001$, not tabulated) after controlling for audit ($F = 1.422, p > .2$, not tabulated) and task experience ($F = 2.574, p > .1$, not tabulated). From the descriptive statistics reported in Tables 2 and 3 and the diagram in Fig. 2 it can be seen that this significant interaction is explained by differences

Table 3. Pension Fund Industry.

<i>Panel A: ANOVA for Pension Fund Industry Performance Level (H_2)</i>				
		<i>F</i>	<i>p</i> ^a	
Specialists versus non-specialists with industry experience		5.215	.026	
Covariates				
Audit experience		0.068	.798	
Industry experience		0.289	.593	
<i>Panel B: Mean (Standard Deviation) [Range]</i>				
Specialists (<i>n</i> = 38)	Non-Specialists with Industry Experience (<i>n</i> = 21)	<i>t</i> ^b	<i>p</i> ^c	
3.55 (1.59) [1.0–7.5]	2.62 (1.76) [0.0–5.5]	2.344	.012	

^aTwo tail.

^bAs Levene's test for equality of variances is not significant ($F = 2.177$, $p > .1$) equal variances are assumed. Shapiro–Wilk test of normality (for sample sizes less than 50) indicates that the industry-specialist ($SW = .971$, $p > .4$) and non-specialist ($SW = .969$, $p > .7$) groups are normally distributed.

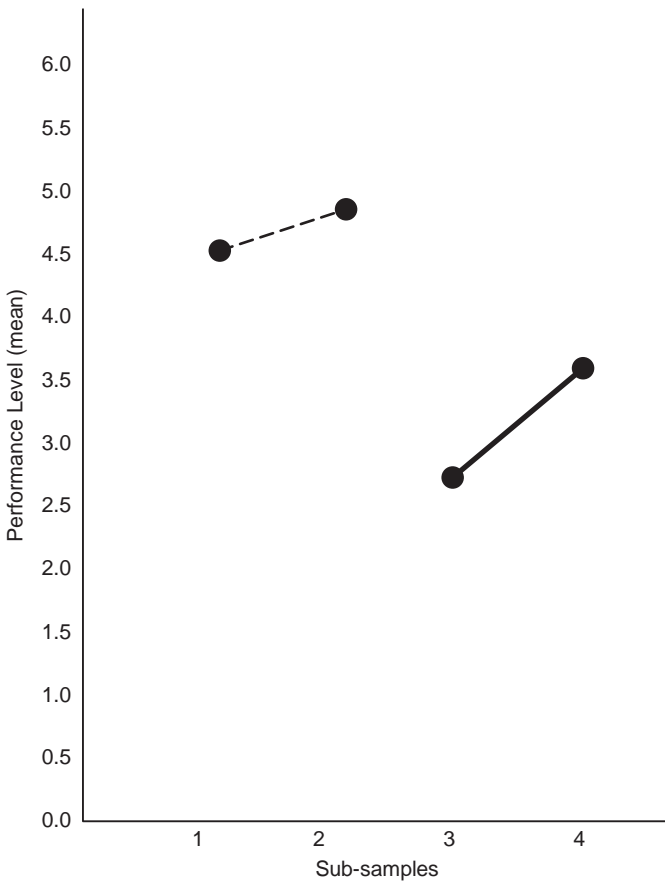
^cOne tail.

between the sub-samples who completed tasks in either the pension fund or manufacturing industries. As hypothesized, the magnitude of the performance differential between industry-specialist and non-specialist auditors is correlated with the level of sub-specialty knowledge required to audit each industry. The difference between industry-specialists and non-specialists with recent experience is significantly greater for a task set in a regulated industry.

Overall, the results indicate that although industry-specialist auditors perform at a higher level than non-specialist auditors with matched industry-based experience, the difference is only significant when performance is dependent on the auditor attaining significant levels of industry-specific sub-specialty knowledge relative to general domain knowledge common to all audits.

SUMMARY AND CONCLUSION

The purpose of this study was to compare and contrast the *relative* performance gains enjoyed by industry-specialist auditors in two very different industry settings, one regulated and the other unregulated. Auditors



Key:

- Sub-sample 1 v Sub-sample 2: Manufacturing Industry
Sub-sample 1 = Non specialists with industry experience
Sub-sample 2 = Industry Specialists

- Sub-sample 3 v Sub-sample 4: Pension Fund Industry
Sub-sample 3 = Non specialists with industry experience
Sub-sample 4 = Industry Specialists

Fig. 2. Diagram of the Difference in Performance Level between Industry-Specialists and Non-Specialists across the Manufacturing and Pension Fund Industries (H_3).

of clients in regulated industries are required to gain relatively more sub-specialty knowledge than auditors of clients in unregulated industries. The acquisition of such sub-specialty knowledge in regulated industries takes time and is gained through extended exposure to clients in the same industry. This difference was hypothesized to result in significant performance gains only in a regulated industry setting when comparing industry-specialist and non-specialist auditors with matched industry-based experience.

An experiment was conducted that compared the task performance of industry-specialists and non-specialist auditors with matched industry-based experience. Pension fund industry-specialist auditors performed significantly better than non-specialist auditors with recent pension fund industry experience. This result is attributed to the significant amount of industry-specific sub-specialty knowledge required to audit clients in a regulated industry. Auditors specializing in regulated industries were found to clearly benefit from greater exposure to clients in the one industry.

In contrast, manufacturing industry-specialists did not perform significantly better than non-specialist auditors with recent manufacturing industry experience. Completing audits in the manufacturing industry typically requires relatively less industry-specific sub-specialty knowledge and non-specialists can acquire that knowledge with relatively limited exposure to clients in the one (unregulated) industry.

The present study contributes to the industry specialization literature by providing evidence that task performance gains enjoyed by industry-specialists are not constant across industries. This finding supports and extends recent research (Dunn & Mayhew, 2004) that has found differences across regulated and unregulated industries, and has implications for how audit firms structure their audit divisions.

The results do not dispute the overarching benefit of industry specialization. Industry-specialists have been found to consistently outperform non-specialists, who have not had the opportunity to gain industry-specific sub-specialty knowledge. But the findings do suggest that care should be taken when using industry specialization as a proxy for audit quality in the economics of auditing research. The criteria used for categorizing firms as industry-specialists can be more relaxed for unregulated industries than for regulated industries, where greater exposure to clients in the one industry setting is clearly beneficial. For example, if it is believed that an audit firm should audit say 20% of the companies in one industry before it can be classified as an industry-specialist firm, the results reported in this chapter suggest that a more lenient, smaller, cut-off can be used for identifying quality auditors of clients in unregulated industries.

From a client perspective, the results suggest that companies in regulated industries can benefit from searching out an industry-specialist audit firm, while those in unregulated industries appear to have the choice between using an industry-specialist firm or another firm that has some other clients in the same industry. From an audit firm perspective, the results suggest that specialization in unregulated industries may not be warranted to achieve the kinds of performance gains normally attributed to all industry-specialists.

The reported results should be considered in light of the normal limitations that apply to experiments. Although this study focuses on one task for each industry, an expert panel was used to ensure the tasks are valid and reflective of real audit tasks the participants are expected to do. Even so, the results are bound by the tasks undertaken. It is possible that the results could differ if a different task was undertaken. Participant co-operation was enhanced by the presence of an experienced researcher for all iterations of the experiment and this study had the full support of senior audit firm personnel. Together these factors increase the probability that the participants completed the task seriously and the task has external validity.

The use of the manufacturing industry as an industry setting is another potential weakness. Clients in this industry span a variety of business types including the manufacture of clothing, steel, motor vehicles and so on. The manufacturing task used in this study has been found to measure manufacturing industry-specialist knowledge (Moroney, 2007) and as such it is valid to be used in the current setting.

Auditor performance is dependant upon an individual's experience, knowledge and ability (Bonner & Lewis, 1990; Libby & Luft, 1993; Libby, 1995). According to Libby (1995) performance is impacted by knowledge directly and by experience indirectly (via knowledge gained from experience). This study uses an individual's experience auditing clients in one industry setting (as either a specialist or a non-specialist with relevant industry experience) to gauge their sub-specialty knowledge. This is a blunt measure and is thus a limitation of the current study.

To obtain industry-specialists and non-specialists with the requisite experience it was necessary to use auditors from different audit firm tiers. Extensive empirical evidence indicates that auditors from first-tier audit firms (industry-specialists) significantly outperform auditors from outside the first-tier (non-specialists) (see e.g., Francis, (2004)). The use of auditors from first- and mid-tier firms biases towards supporting a significant difference for the regulated industry (H_2) and biases against finding support for an insignificant difference for the unregulated industry (H_1) or for a difference in behavior for the two industry settings (H_3). The test comparing

the performance of all auditors across the tiers reported in the data analysis section was not significant, indicating that between tier differences do not explain the results presented here.

This study is the first to provide evidence of the relative performance gains available to industry-specialist auditors in different industry settings when compared with non-specialist auditors with recent industry-based experience. Future research could extend this comparison to different industries and/or attributes that vary across industries and influence the demand for industry-specialist auditors, such as research and development intensity (Godfrey & Hamilton, 2005).

NOTES

1. The manufacturing industry was selected as the unregulated industry and the pension fund industry selected as the regulated industry with the aid of the participating first-tier firms on the basis of the stark contrast between these industries. While there is variety within the manufacturing industry (e.g., motor vehicle, steel and clothing), the task used was developed by first-tier manufacturing industry-specialist partners with a view to be as relevant as practicable across the manufacturing industry. This is not the first study to use the manufacturing industry as a context when assessing the benefits of industry knowledge (e.g., Bedard & Biggs, 1991a, 1991b).

2. The first-tier firms participated in the first iteration of the experiment in December 2002 and the mid-tier firms took part in the second iteration in November 2004. Partners from the expert panel, who helped develop the tasks, confirmed that there had been no change in the intervening two years that would have prevented use of the tasks, information cues or solutions.

3. All participants completed both tasks. The order of the two tasks was varied between subjects to control for any demand effects. The order the tasks were completed in did not impact performance ($F = 1.777, p = .183$).

4. These tasks were used in Moroney (2007) to assess industry-specialist efficiency and effectiveness. Moroney (2007) found that both groups of industry-specialists performed significantly better when working in specialization than when working out of specialization. Thus, these tasks test industry sub-speciality knowledge.

5. Each task came with information cues, which included excerpts from accounting standards and industry regulation. The cues did not add any new information to the basic task material.

6. Inter-rater agreement was measured using Cohen's Kappa (1960). The Kappa for the manufacturing task was 0.9424 and 0.9113 for the pension fund task (all are significant at $p < .001$). This means that over 91% of the joint agreements were agreements, with chance excluded.

7. An additional 11 first-tier and 2 mid-tier auditors with less than 2 years experience who completed the experiment were excluded from the analysis.

8. Because relatively few non-specialist auditors had pension fund industry experience, the 12 auditors were included in the pension fund industry group.

9. Sensitivity tests excluding individuals who reported no task experience are reported later in the chapter.

10. This observation was confirmed by the audit partners who provided access to the participants.

11. A Kruskal–Wallis test was run to compare performance across the 12 participating audit firms. The result is not significant ($\chi^2 = 15.087$, $df = 11$, $p > .1$, not tabulated). This suggests that the populations from which the groups are sampled do not significantly differ.

12. This between tier result holds for auditors working on a task where they have industry-based experience ($F = 1.469$, $p > .2$, not tabulated) as well as for a task where they do not have industry-based experience ($F = .969$, $p > .3$, not tabulated).

13. Less than 50% of the participants undertaking the manufacturing task reported having task experience. The participants were split into two groups and H_1 was retested. The mean performance level for industry-specialist (non-specialist) auditors who reported no task experience is 4.83 (4.02) (not tabulated), the differences is not significant ($t = 1.592$, $p > .1$, not tabulated). Industry-specialist (non-specialist) auditors with task experience achieved a mean performance level of 5.09 (5.06) (not tabulated), the difference is not significant ($t = .010$, $p > .9$, not tabulated). Thus, task experience, or lack of it, does not appear to drive the results reported here.

14. The theoretical range for both cases is 0–10.

15. The Levene's test for equality of variances across all four groups is not significant ($F = 1.507$, $p > .2$); homogeneity of variance is therefore assumed. Each sub-sample is normally distributed. Thus, the sample does not violate the assumptions for conducting an ANOVA.

ACKNOWLEDGMENTS

This chapter has benefited from the helpful comments of Vicky Arnold, editor, the associate editor and two reviewers, Jere Francis and Ken Trotman and participants at the American Accounting Association Annual Meeting, Washington, DC and the Accounting and Finance Association of Australia and New Zealand Conference, Wellington, as well as participants at Monash University and RMIT research seminars. We gratefully acknowledge the generous support of the audit firms who participated in this study. In particular, we thank the expert panel that helped write the tasks and model solutions and the participants for their time.

REFERENCES

- Abdolmohammadi, M. J., Seafoss, D. G., & Shanteau, J. (2004). An investigation of the attributes of top industry audit specialists. *Behavioral Research in Accounting*, 16, 1–17.

- Balsam, S., Krishnan, J., & Yang, J. (2003). Auditor industry specialization and earnings quality. *Auditing: A Journal of Practice and Theory*, 22, 71–97.
- Bedard, J., & Biggs, S. (1991a). Pattern recognition, hypotheses generation and auditor performance in an analytical task. *The Accounting Review*, 66, 622–642.
- Bedard, J., & Biggs, S. (1991b). The effect of domain-specific experience on evaluation of management representations in analytical procedures. *Auditing: A Journal of Practice and Theory*, 10(Suppl.), 77–90.
- Bédard, J. (1989). Expertise in auditing: Myth or reality? *Accounting, Organizations and Society*, 14, 113–131.
- Bédard, J., & Chi, M. T. H. (1993). Expertise in auditing. *Auditing: A Journal of Practice and Theory*, 12(Suppl.), 21–45.
- Biggs, S. F., Selfridge, M., & Krupka, G. R. (1993). A computational model of auditor knowledge and reasoning processes in the going-concern judgment. *Auditing: A Journal of Practice and Theory*, 12(Suppl.), 82–99.
- Bonner, S. (1990). Experience effects in auditing: The role of task-specific knowledge. *The Accounting Review*, 65, 72–92.
- Bonner, S. E., & Lewis, B. L. (1990). Determinants of auditor expertise. *Journal of Accounting Research*, 28(Suppl.), 1–19.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Craswell, A. T., Francis, J. R., & Taylor, S. (1995). Auditor brand name reputations and industry specializations. *Journal of Accounting and Economics*, 20, 297–322.
- DeFond, M. L., Francis, J. R., & Wong, J. (2000). Auditor industry specialization and market segmentation: Evidence from Hong Kong. *Auditing: A Journal of Practice and Theory*, 19(1), 49–66.
- Dunn, K. A., & Mayhew, B. A. (2004). Audit firm industry specialization and client disclosure quality. *Review of Accounting Studies*, 9, 35–58.
- Francis, J. R. (2004). What do we know about audit quality? *The British Accounting Review*, 36(1), 345–368.
- Francis, J. R., Reichelt, K., & Wang, D. (2005). The pricing of national and city-specific reputations for industry expertise in the US audit market. *The Accounting Review*, 80(1), 113–136.
- Francis, J. R., Stokes, D. J., & Anderson, D. (1999). City markets as a unit of analysis in audit research and the re-examination of Big 6 market shares. *Abacus*, 35(2), 185–206.
- Godfrey, J. M., & Hamilton, J. (2005). The impact of R&D intensity on demand for specialist auditor services. *Contemporary Accounting Research*, 22(1), 55–93.
- Hammersley, J. S. (2006). Pattern identification and industry-specialist auditors. *The Accounting Review*, 81, 309–336.
- Hogan, C., & Jeter, D. (1999). Industry specialization by auditors. *Auditing: A Journal of Practice and Theory*, 18(1), 1–17.
- Krishnan, G. V. (2003). Does Big 6 auditor industry expertise constrain earnings management? *Accounting Horizons*, 17, 1–16.
- Kwon, S. (1996). The impact of competition within the client's industry on the auditor selection decision. *Auditing: A Journal of Practice and Theory*, 15(1), 53–69.
- Libby, R. (1995). The role of knowledge and memory in audit judgment. In: R. H. Ashton & A. H. Ashton (Eds), *Judgment and decision making research in accounting and auditing* (pp. 176–206). Cambridge: Cambridge University Press.

- Libby, R., & Luft, J. (1993). Determinants of judgment performance in accounting settings: Ability, knowledge, motivation and environment. *Accounting, Organizations and Society*, 18, 425–450.
- Libby, R., & Tan, H. T. (1994). Modeling the determinants of audit expertise. *Accounting, Organizations and Society*, 19(8), 701–716.
- Low, K. Y. (2004). The effects of industry specialization on audit risk assessments and audit-planning decisions. *The Accounting Review*, 79, 201–219.
- Marchant, G. (1990). Discussion of determinants of auditor expertise. *Journal of Accounting Research*, 28, 21–28.
- Moroney, R. (2007). Does industry expertise improve the efficiency of auditor judgment? *Auditing: A Journal of Practice and Theory*, 26, 69–94.
- Owhoso, V. E., Messier, W. F., & Lynch, J. G. (2002). Error detection by industry-specialized teams during sequential audit review. *Journal of Accounting Research*, 40, 883–900.
- Simnett, R., Lockett, P., & Wright, A. M. (2000). The development of specialists' industry knowledge in government auditing. *Australian Accounting Review*, 10, 35–45.
- Solomon, I., Shields, M. D., & Whittington, O. R. (1999). What do industry-specialist auditors know? *Journal of Accounting Research*, 37, 191–208.
- Spence, M. T., & Brucks, M. (1997). The moderating effects of problem characteristics on experts' and novices' judgments. *Journal of Marketing Research*, 34, 233–247.
- Taylor, M. H. (2000). The effects of industry specialization on auditors' inherent risk assessments and confidence judgments. *Contemporary Accounting Research*, 17, 693–712.
- Thibodeau, J. C. (2003). The development and transferability of task knowledge. *Auditing: A Journal of Practice and Theory*, 22, 47–68.
- Wright, S., & Wright, A. M. (1997). The effect of industry experience on hypothesis generation and audit planning decisions. *Behavioral Research in Accounting*, 9, 273–294.

AUDITORS' SELF-OTHER AGREEMENT ON PERCEIVED POSSESSION OF EXPERT ATTRIBUTES

Mohammad J. Abdolmohammadi

ABSTRACT

In this study 43 auditors of varying rank (staff/assistant, senior/supervisors, and managers/partners) and expertise level (candidates for specialty, competent specialists, and expert specialists) assessed the degree to which they believed themselves and their colleagues possessed detailed expert attributes. Definitions of 11 attributes that were found by Abdolmohammadi, Searfoss, and Shanteau (2004) to be extremely or very important to expertise in audit specialty were provided to the subjects for their assessment. As hypothesized, the possession of many attributes that can be classified as trainable and developable differed by professional rank. However, innate attributes of intelligence and quick thinker did not differ by professional rank. Also, as hypothesized, systematic biases in assessment of possession of attributes of superiors and subordinates were observed, as well as evidence of inflated bias of self by some participants. Implications for accounting practice, education, and research are discussed.

Advances in Accounting Behavioral Research, Volume 11, 81–102

Copyright © 2008 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11004-3

INTRODUCTION

The primary objective of this chapter is to provide evidence on auditors' perceived possession of detailed expert attributes by themselves and their colleagues of varying rank (staff/assistant, senior/supervisors, and managers/partners) and expertise level in one's professional specialty (candidates for specialty, competent specialists, and expert specialists). Understanding of self-other perceptions of expert attributes is important from an organizational/individual outcomes perspective. While high agreement in perceptions of attribute possession may result in positive attitude toward work (an enhanced organizational/individual outcome), low agreement may increase dissatisfaction and turnover intentions (cf., [Chatman, 1991](#)). [Yammarino and Atwater \(1993\)](#) argue that when self-other perceptions agree, an accurate estimation of possession of expert attributes will be indicated, which in turn will result in enhanced organizational/individual outcomes. However, if self-other perceptions indicate an overestimation bias, then a diminished organizational/individual outcome will result. Conversely, if self-other perceptions result in an underestimation bias, then a mixed organizational/individual outcome may result.

The organizational/individual outcomes perspective has important implications for the audit setting because the literature indicates that auditors engage in a strategic behavior game in which they try to understand the general perceptions and expectations of their competence by their colleagues. As [Tan and Jamal \(2006, p. 765\)](#) note, "In a strategic game, each auditor must form expectations about the knowledge and behavior of another party [client management or another auditor] in order to decide what action to take." Consequently, auditors' perceptions about what other auditors know about them are an important factor in their strategic behavior game ([Rich, Solomon, & Trotman, 1997](#); [Gibbins & Trotman, 2002](#); [Tan & Jamal, 2006](#)). These findings suggest that professional auditors have interest in general perceptions of other auditors about their possession of detailed expert attributes to manage their own reputation. The social psychology literature (e.g., [Depaulo, Kenny, Hoover, Webb, & Oliver, 1987](#)) indicates that management of others' impressions requires one to have accurate perception of what others know about him/her, and the auditing literature (e.g., [King, 1996](#); [Mayhew, 2001](#)) provides empirical evidence that practicing auditors do try to enhance their reputation by managing the perceptions of their colleagues about their abilities.

The current study is also important from a professional responsibility perspective. Rule 201 of the AICPA's Code of Professional Conduct

requires auditors to have professional competence (i.e., “undertake only those professional services that the member or the member’s firm can reasonably expect to be completed with professional competence”) (AICPA, 1988). Surprisingly, very little evidence is reported in the literature on the level at which auditors of varying expertise and professional ranks possess various attributes of competence. As Stone, Hunton, and Wier (2000, p. 697) argue, organizations, including accounting firms, will benefit from understanding of the available knowledge and skills as well as potentially unmet knowledge needs. These authors provide evidence of differences in knowledge and ability across three levels of professional management accounting ranks such as juniors, seniors, and managers. The current study provides perceptual data by professional rank and expertise level in public accounting.

The remainder of the chapter is organized as follows. The background literature is presented in the next section as a means of developing the study’s research hypotheses. The research method and data analysis are presented in the subsequent two sections, followed by a summary of the findings and several implications in the final section.

BACKGROUND AND HYPOTHESES

Possession of Expert Attributes

Early literature on expertise in auditing focused on technical knowledge and problem-solving ability (e.g., Bonner & Lewis, 1990). An outcome of this literature has been the formulation of performance as a function of attributes such as ability, knowledge, motivation, and environment by Libby and Luft (1993).¹ A more recent focus of this literature has been on identification of detailed attributes that are important to expertise in auditing (cf., Abdolmohammadi & Shanteau, 1992; Libby & Tan, 1994; Tan & Libby, 1997; Tan, 1999; Abdolmohammadi, Searfoss, & Shanteau, 2004). The detailed expertise attributes include knowledge, experience, and problem-solving ability that have been investigated extensively in prior literature as well as a number of attributes such as confidence, communication skills, and research skills that have not been extensively studied in prior research. Typically, the studies of attribute importance have asked auditors to rank-order various attributes by their importance to audit expertise. For example, a sample of 114 partners rated as top industry audit specialists in Abdolmohammadi et al. (2004) rank-ordered

25 expert attributes from 1 to 5 (minimally important to extremely important). The subjects in the current study assessed the possession of all attributes in [Abdolmohammadi et al.'s \(2004\)](#) study. However, only 11 attributes were rated by [Abdolmohammadi et al. \(2004\)](#) as extremely important or very important to audit expertise. These attributes are the focus of the current chapter.² The appendix lists these attributes and their definitions. Also included in the appendix are the percentages of [Abdolmohammadi et al.'s \(2004\)](#) participants who judged each attribute as trainable, developable, or innate.³ As reported in the Additional Analysis section, these results can explain differences in attribute possession by rank.

Professional rank has been investigated in studies of relative importance (but not level of possession) of expert attributes in the past (e.g., [Bhamornsiri & Guinn, 1991](#); [Emby & Etherington, 1996](#); [Tan & Libby, 1997](#); [Tan, 1999](#)). For example, [Bhamornsiri and Guinn \(1991\)](#) investigated the relative importance of some attributes for promotion to higher professional ranks. The authors found technical competence to be the most important attribute for promotion from staff to senior and from senior to manager, but not from manager to partner level. For the latter group the most important attribute was communication skills followed by interpersonal skills, then technical skills, practice development, and administrative skills. Communication and interpersonal skills were less important to the promotion decision from senior to manager, and to even a lesser degree for promotion of staff to senior rank. Practice development and administrative skills were not important for the promotion from staff to senior, but moderately important for promotion of seniors to manager level. Similarly, [Tan and Libby \(1997\)](#) found that while tacit managerial knowledge was important for superior performance at the manager level, technical skills and problem-solving abilities were, respectively, more important for superior performance at the staff and senior rank levels.⁴ Finally, [Bierstaker and Wright \(2001\)](#) find that experience and problem-solving ability are important for superior performance at the senior rank.

Collectively, studies of expert attributes show differences by professional rank. In particular, experienced auditors generally have accurate perceptions of their own technical competence ([McKnight & Wright, 2005](#)) and what their superiors (e.g., managers in [Tan & Jamal, 2006](#)) think of their technical competence. The current chapter extends this literature by presenting data on the degree to which auditors of varying ranks perceive their own and their colleagues' possession of various expert attributes.

Based on the general expectation from prior literature, the study's first hypothesis is stated as:

H₁. The perceived possession of expert attributes increases by professional rank.

Self-Other Agreement on Possession of Expert Attributes

McKnight and Wright (2005) find that auditors generally understand the relative weights of attributes that lead to various performance levels and are capable of assessing the degree to which they possess various attributes. The broader organizational theory literature suggests that assessment of attribute possession of self and others is a reliable methodology (Campbell & Lee, 1988; Litras, 1980) and a positive aspect of performance evaluation (Lawrie, 1989). For example, in a study of Taiwanese managers, Silverthorne (2000) found that there is generally a good match between self-perception of leaders and perceptions by peers, superiors, and subordinates.⁵ Atwater, Ostroff, Yammarino, and Fleenor (1998) present similar results from a large sample of mid-level managers who participated in a leadership development program in the United States. However, this literature also presents evidence of a bias indicating that self-other perceptions are not always calibrated. Thus, Atwater et al. (1998) recommend simultaneous consideration of self-ratings and other ratings in explaining managerial effectiveness because there are differences in possession perceptions between these self-peer ratings.

The organizational theory literature also indicates that self-other "ratings or evaluations may address any of the areas relevant for personnel and human resource management, such as *personality characteristics, individual abilities/skills*, training needs assessments, job analysis ratings, performance appraisals, or leadership behaviors" Yammarino and Atwater (1993, p. 232, emphasis added as they are related to the current study). Expert audit attributes investigated in the current study have personality characteristics (e.g., self-confidence) as well as individual abilities (e.g., problem solver) and skills (e.g., knowing what is relevant).

An important issue of research in this literature has been the level of agreement between assessors. In a meta analysis of this literature, Harris and Schaubroeck (1988) concluded that while the correlations between peer and supervisor ratings of an individual are relatively high, the correlations between self-supervisor and self-peer ratings are only moderate (i.e., there are differences). Harris and Schaubroeck's (1988) review also indicates that

the convergence between self-peer and self-supervisor ratings is particularly low for managerial/professional staff (see also Bass & Yammarino, 1991; Furnham & Stringfield, 1994), a reference group that is similar to the auditor-subjects in the current study. As summarized in Fig. 1, low agreement has undesirable effects on organizational or individual outcomes. For example, Atwater et al. (1998) report that managerial effectiveness is lowest for overassessors when self-ratings are moderate and subordinate ratings are low, and that managerial effectiveness tends to increase for underassessors and decrease for overassessors. On the other hand, if self-other perceptions agree, an accurate estimation of possession of expert attributes will be indicated, which in turn will result in enhanced organizational/individual outcomes.

In conclusion, due to over or underassessment biases, self-other attribute assessments are not always calibrated, and the non-auditing literature is mixed on the level and direction (i.e., over or underassessment) of this bias (cf., Furnham & Stringfield, 1994, p. 62). Investigation of this bias in auditing is desirable because evidence on it in auditing is very limited and yet it is mixed.⁶ Specifically, while Kennedy and Peecher (1997) find that auditors are overconfident in assessing their own and their subordinates' technical knowledge, other studies show that managers and partners have the highest level of accuracy in their perceptions of their own performance (McKnight & Wright, 2005) and meta perception of what others know of their technical competence (Tan & Jamal, 2006). The current chapter compares assessments of specialty candidates (i.e., novices) and competent

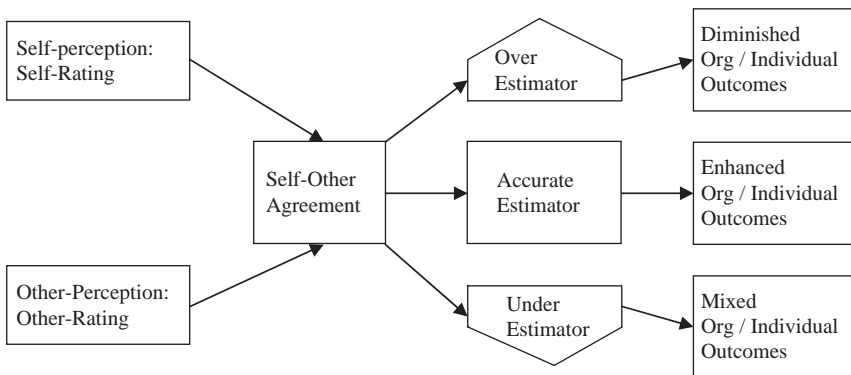


Fig. 1. A Model of Self-Perception Accuracy. Source: Yammarino and Atwater (1993).

specialists (i.e., intermediate specialists) to those of expert specialists as a criterion group. The results add to the very limited and growing body of literature in auditing. In particular, the focus of this chapter is on differences by expertise in audit specialty. Due to the mixed results in the literature regarding the direction of the bias, the following two hypotheses are presented in a bi-directional mode:

H_{2a}. Specialty candidates' perceptions of attribute possession of others (i.e., superiors) are different from those of expert specialists.

H_{2b}. Competent specialists' assessments of attribute possession of others (i.e., superiors and subordinates) are different from those of expert specialists.

Finally, the organizational theory literature suggests that when participants assess their own attributes as well as the attributes of their peers, they tend to have a bias toward inflating their own ratings (Yammarino & Atwater, 1993). This finding suggests the final hypothesis as follows:

H₃. Auditors' assess the possession of expert attributes of their peers at lower levels than their own.

RESEARCH METHOD

The 11 attributes rated as extremely important or very important to auditing expertise (see the appendix) are investigated for possession by self and other auditors. Eighty-five packets were sent to contact partners in five offices of international accounting firms and a regional firm in the Northeastern United States. The partners were requested to contact auditors of varying ranks and expertise in their specialties for participation in the study. The contact partners returned 43 responses yielding a response rate of 50.6%. No follow up request for data collection was sent. The research packet had a cover page in which the purpose of the study was described and definitions were provided for three generic levels of specialty as follows⁷:

Specialty candidate (novice)	One who possesses the characteristics necessary to be admitted for training in an audit specialty
Competent specialist (intermediate)	One who is trained in an audit specialty area and has also practiced for 1–2 years in the specialty
Expert specialist (expert)	One who has mastered the tasks and has become an expert in the specialty

Table 1. Demographic Information.

Category	N	Experience		Age	
		Mean	Standard deviation	Mean	Standard deviation
<i>Specialty level</i>					
Candidate	13	1.00	0.87	23.31	2.66
Competent	18	2.61	1.38	25.94	2.58
Expert	8	5.38	3.07	36.29	9.30
Aggregate	39 ^a	2.73	2.33	27.14	6.26
<i>Professional rank</i>					
Staff/Assistant	16	0.89	0.65	23.13	2.31
Senior/Supervisor	13	2.68	0.78	26.62	2.36
Manager/Partner	13	4.92	2.68	33.33	8.14
Aggregate	42 ^a	2.73	2.33	27.14	6.26

^aDifference due to missing data.

In a demographic questionnaire, the participants provided information about their primary areas of audit specialization and their level of expertise in their specialty, years of experience in the specialty, and their professional rank (i.e., staff, assistant, senior, supervisor, manager, or partner). Table 1 presents summary statistics on experience and age of the participants by level of expertise and professional rank. Of the 43 participants one did not reveal his/her rank and four did not reveal their level of expertise in their specialty, reducing the usable data for analysis by rank to 42 and by level of expertise in specialty to 39.

As Table 1 shows, of the 39 participants (10 females and 29 males) who identified their specialty level, 13 were specialty candidates who on average had one year of audit experience and 23.31 years of age. Competent specialists were 18 professionals who on average had 2.61 years of audit experience and 25.94 years of age. The remaining 8 participants were experts in their specialty who on average had 5.38 years of experience, and 36.29 years of age. Classification by professional rank resulted in 16 staff/assistants, 13 senior/supervisors and 13 manager/partners whose experience, age, and gender are specified in Table 1.

Other demographic information (not tabulated) indicated that while 24% of the participants had graduate degrees, 76% possessed bachelor's degrees in accounting or a related field. Also, approximately one-half of the participants (19) specialized in high technology or financial services, while the remaining-half (20) had specialties in nine other areas such as insurance, healthcare, or real estate.

The remainder of the task instrument had two sections. Section I provided the listing and definitions (but not the importance ratings) of the 11 pre-defined expert attributes as listed in the appendix. In Section II, the participants were provided a questionnaire in which they were asked to assess the degree to which they personally possessed each of the attributes. Using the levels of expertise and professional rank from the demographic questionnaire, this data serve to investigate differences in attribute possession by auditors of varying expertise level and professional rank. The participants were also asked to think of a specialty candidate, a competent specialist, and an expert specialist that they personally knew and then assess the degree to which the participants believed specialty candidates, competent specialists, and expert specialists possessed each of the 11 attributes. The scale used for Section II was a Likert 1–5 scale (minimal possession to very high possession, the mid-point being moderate possession).

DATA ANALYSIS

Perceived Attribute Possession by Professional Rank (H_1)

H_1 predicts that the perceived possession of expert attributes *increase* by professional rank. The non-parametric multi-sample Kruskal–Wallis test was performed on each of the 11 expert attributes to investigate differences by professional rank. The results that are reported in [Table 2](#) are consistent with H_1 . Specifically, of the 11 attributes, nine (i.e., the first nine listed in [Table 2](#)) show statistical significance at 0.024 or better. The remaining two, Intelligence and Quick Thinker, are not significant. Comparing these results with the appendix shows an interesting pattern: for attributes that [Abdolmohammadi et al.'s \(2004\)](#) subjects rated primarily as trainable and/or developable; there is a significant rank effect. However, for the two that the subjects rated as innate (Intelligence and Quick Thinker), there is no rank effect.

Perceptions of Expert Attribute Possessions by Others (H_{2a} and H_{2b})

H_{2a} and H_{2b} predicts that specialty candidates' (competent specialists') perceptions of attribute possession of others are different from those of expert specialists. To test for these hypotheses, the aggregate attribute assessments of all 11 attributes are compared by levels of specialty.

Table 2. Perceptions of Personal Possession of Attributes by Professional Rank – Scale: 1–5 (Minimal to Extreme Possession).

Attribute (Per Appendix)			Mean	Median	Standard Deviation
1	Current knowledge	Assistant/Staff	2.75	3.00	0.86
		Senior/Supervisor	3.53	4.00	0.52
		Manager/Partner	4.23	4.00	0.73
Kruskal–Wallis <i>H</i> -statistic (significance)				18.13 (<0.001)	
2	Problem solver	Assistant/Staff	3.44	3.00	0.63
		Senior/Supervisor	3.46	3.00	0.66
		Manager/Partner	4.31	4.00	0.63
Kruskal–Wallis <i>H</i> -statistic (significance)				12.16 (0.002)	
3	Experience	Assistant/Staff	2.75	3.00	1.00
		Senior/Supervisor	3.62	4.00	0.65
		Manager/Partner	4.42	4.00	0.52
Kruskal–Wallis <i>H</i> -statistic (significance)				20.17 (<0.001)	
4	Perceptive	Assistant/Staff	3.50	3.50	0.73
		Senior/Supervisor	3.62	4.00	0.51
		Manager/Partner	4.23	4.00	0.73
Kruskal–Wallis <i>H</i> -statistic (significance)				7.49 (0.024)	
5	Communicates expertise	Assistant/Staff	2.75	3.00	0.78
		Senior/Supervisor	3.46	4.00	0.66
		Manager/Partner	4.39	4.00	0.65
Kruskal–Wallis <i>H</i> -statistic (significance)				21.35 (<0.001)	
6	Self-confidence	Assistant/Staff	3.38	3.00	0.62
		Senior/Supervisor	3.85	4.00	0.56
		Manager/Partner	4.15	4.00	0.56
Kruskal–Wallis <i>H</i> -statistic (significance)				10.40 (0.006)	
7	Adaptability	Assistant/Staff	3.50	3.50	0.52
		Senior/Supervisor	3.62	4.00	0.51
		Manager/Partner	4.15	4.00	0.70
Kruskal–Wallis <i>H</i> -statistic (significance)				7.47 (0.024)	
9	Knows what is relevant	Assistant/Staff	3.13	3.00	1.03
		Senior/Supervisor	3.62	4.00	0.51
		Manager/Partner	4.31	4.00	0.63
Kruskal–Wallis <i>H</i> -statistic (significance)				12.50 (0.002)	
10	Assumes responsibility	Assistant/Staff	2.63	4.00	0.81
		Senior/Supervisor	4.00	4.00	0.58
		Manager/Partner	4.54	5.00	0.52
Kruskal–Wallis <i>H</i> -statistic (significance)				10.79 (0.005)	
8	Intelligence	Assistant/Staff	3.69	4.00	1.00
		Senior/Supervisor	3.69	4.00	0.65
		Manager/Partner	4.00	4.00	0.52
Kruskal–Wallis <i>H</i> -statistic (significance)				2.20 (0.191)	
11	Quick thinker	Assistant/Staff	3.38	3.00	0.89
		Senior/Supervisor	3.54	4.00	0.66
		Manager/Partner	3.92	4.00	0.76
Kruskal–Wallis <i>H</i> -statistic (significance)				4.42 (0.110)	

Note: Significance levels are adjusted for ties. Assistant/Staff = A; Senior/Supervisor = B; Manager/Partner = C.

Table 3. Aggregate Results of Attribute Possession Assessment of Others – Scale: 1–5 (Minimal Possession to Very High Possession).

Person Assessing	<i>N</i>	Mean	Median	Standard Deviation	Average Rank	Z-Statistic (Significance)
<i>Panel A: Perceived attribute possession of specialty candidates</i>						
Specialty candidate	143	3.20	3.00	0.83	256.60	4.91 (<0.001)
Competent specialist	198	2.58	3.00	1.00	183.70	-4.85 (<0.001)
Expert specialist	88	2.88	3.00	0.86	217.90	0.25 (0.401)
Overall rank					215.00	
Kruskal-Wallis- <i>H</i> (significance)			28.80 (<0.001)			
<i>Panel B: Perceived attribute possession of competent specialists</i>						
Specialty candidate	143	4.00	4.00	0.68	257.40	5.01 (<0.001)
Competent specialist	198	3.55	4.00	0.74	188.50	-4.10 (<0.001)
Expert specialist	88	3.68	4.00	0.65	205.80	-0.78 (0.218)
Overall rank					215.00	
Kruskal-Wallis- <i>H</i> (significance)			26.24 (<0.001)			
<i>Panel C: Perceived attribute possession of expert specialists</i>						
Specialty candidate	143	4.55	5.00	0.71	229.30	1.69 (0.046)
Competent specialist	198	4.39	5.00	0.74	201.20	-2.13 (0.017)
Expert specialist	88	4.52	5.00	0.69	222.70	0.66 (0.255)
Overall rank					215.00	
Kruskal-Wallis- <i>H</i> (significance)			4.69 (0.096)			

The results are presented in Table 3 where the non-parametric Kruskal-Wallis test is used for each of the specialty groups to analyze assessment differences. Also presented are the Z-statistics that indicate the differences between each specialty level and the overall rank in each panel.

Table 3 shows the Kruskal-Wallis *H*-statistics for specialty candidates (Panel A), competent specialists (Panel B), and expert specialist (Panel C). The results are highly significant ($p < 0.001$) for panels A and B, but marginally significant (H -statistic = 4.69, $p = 0.096$) for Panel C. In all three panels, however, the Z-statistics indicate that, competent specialists' perceived possession of attributes were at lower levels than those of expert specialists (H_{2b}). Also, in support of H_{2a} , specialty candidates assessed the levels of attribute possession of others at significantly higher levels than those of expert specialists. The data in panels A, B, and C clearly indicate that regardless of the level of expertise of the assessing group, attribute possession is positively associated with levels of expertise.

*Comparison between Perceived Possession of Attributes
by Self and Peers (H_3)*

H_3 predicts that auditors will rate their own possession of expert attributes at higher levels than their peers. The Mann–Whitney tests of the aggregate ratings of all 11 attributes for the three specialty levels comparing the attribute possession assessments of self and peers are summarized in Table 4. The results for competent specialists provide support for H_3 . Specifically, competent specialists assessed their own possession (mean = 3.83) at higher levels than those of their peers (mean = 3.55) and the difference was highly significant (W -statistic = 43,151.0, $p < 0.001$). However, specialty candidates' assessments did not indicate significant differences, and thus do not provide support for H_3 . Also, the difference between self and peer possessions is highly significant for the expert specialists (W -statistic = 6,739.5, $p = 0.006$), but in an opposite direction to H_3 . Specifically, expert specialists assessed attribute possession of other experts (mean = 4.52) at higher levels than themselves (mean = 4.24). These results provide mixed evidence on H_3 . The implications of these results are discussed in the final section.⁸

SUMMARY AND CONCLUSIONS

Auditors with varying degrees of expertise in their audit specialty (specialty candidates, competent specialists, and expert specialists) and professional rank (staff/assistant, senior/supervisors, and managers/partners) from several accounting firms assessed the degree to which they personally possessed each of 11 pre-defined attributes (see the appendix). The participants also assessed the degree of possession of these attributes by other specialists of varying expertise. The results indicate that perceived possession of attributes classified as trainable and/or developable increases by professional rank. Similar to prior research (e.g., Tan & Libby, 1997; Tan, 1999; Wright, 2001; Abdolmohammadi et al., 2004) that has found importance of expertise attributes to increase by professional rank, these results show that they are also perceived to be gaining more possession as one moves up the ladder of professional rank. However, attributes classified as innate (intelligence and quick thinker) did not differ significantly by professional rank.

The next important finding is that while in comparison to expert specialists, specialty candidates assess attribute possessions of others at

Table 4. Perceived Attribute Possession of Self-Other – Scale: 1–5 (Minimal Possession to Very High Possession).

Person Assessing	Self			Other									Mann–Whitney <i>W</i> -statistic (significance)
				Candidates (<i>n</i> = 143)			Competents (<i>n</i> = 198)			Experts (<i>n</i> = 88)			
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	Mean	Median	Standard deviation	Mean	Median	Standard deviation	
Candidate	3.18	3.00	0.79	3.20	3.00	0.83							20,529.0 (0.991)
Competent	3.83	4.00	0.69				3.55	4.00	0.74				43,151.0 (< 0.001)
Expert	4.24	4.00	0.69							4.52	5.00	0.69	6,739.5 (0.006)

a significantly higher level, competent specialists systematically perceive others at lower levels of attribute possession. The disagreement between lower levels of specialty with respect to attribute possession may result in loss of calibration in performance evaluations of subordinates. Finally, a comparison was made between the assessment of attribute possession by self and peers (i.e., same specialty level). This analysis indicates that while competent specialists (expert specialists) assessed their own possession of attributes at higher (lower) levels than their peers, specialty candidates' self-other assessments did not indicate significant differences.

Implications

The results of this study may be useful for designing training material for specialty attributes. The innate attributes, the possession of which may not differ by expertise level (e.g., intelligence and quick thinker) may be used for screening purposes at the time of hiring. This result counters the general assumption that only those with superior levels of intelligence and quick thinking are promoted to higher ranks in public accounting. Thus, the finding of insignificant differences by professional rank for innate attributes may need further investigation.

The trainable and/or developable attributes that showed significant possession differences by professional rank may require differential training by rank. Accounting firms can also use levels of attribute possession as a guide for task assignment and performance evaluation. In particular, assignment of complex tasks requires special care in selecting auditors who possess the right level of expert attributes to perform the task. The literature indicates that auditor performance for complex tasks is lower in effectiveness than simple tasks (see [Bonner, 1994](#) for a review) and that the accuracy of knowing what other auditors think of one's technical proficiency decreases with increases in task complexity ([Tan & Jamal, 2006](#)). In other words, task complexity moderates the effects of the accuracy of knowing what other auditors think of one's technical proficiency. These results indicate that the relationship between levels of task complexity and attribute possession await investigation in the future. These results also may imply that auditors should rely on the work of expert specialists to perform complex tasks but that they tend to under-rely on these expert specialists, a possibility that is worthy of investigation in prior research.

Behavioral research can also benefit from the results of the current study by devising experiments in which exact nature and measurement of various attributes can be investigated. For example, a repository of possession of "current knowledge" can be developed by expertise level. Such a knowledge base can help pinpoint the exact nature of current knowledge differences by expertise level. These assessment methods can then be used to investigate their relationship with superior performance. While some studies of this type have been conducted (e.g., Bonner & Lewis, 1990; Tan & Libby, 1997), the number of attributes investigated has been limited (e.g., knowledge and innate ability in Bonner & Lewis, 1990) and the assessment methods have been fairly generic (e.g., the use of Graduate Record Examination questions for measuring general knowledge in Bonner & Lewis, 1990). Studies of a larger number of attributes and more specific knowledge determinants may be needed to better assess expertise differentials in various specialty areas. Future behavioral research could also examine how differences in perceptions about expert attributes could result in diminished auditor communication or performance.

Accounting education can also benefit from consulting the results of this study. In general the focus of much of accounting education is on the development of students' technical knowledge and problem-solving skills. The results of this study indicate that attention to developing students' communication skills, self-confidence, adaptability, and other attributes that show differential degrees of possession by auditors of varying rank and expertise may also be warranted. Also, while innate attributes (e.g., intelligence and quick thinker) may be used as screening devices for admission to majors (e.g., accounting) and course (e.g., auditing), other attributes such as communication skills can be further emphasized in accounting courses to better prepare students for entry to the accounting profession.

Investigation of differences by expertise level indicated that in comparison to expert specialists, specialty candidates generally overassessed others, while competent specialists systematically underassessed others. This finding indicates that there is disagreement between lower levels of specialty with respect to attribute possession. For competent specialists this finding indicates that they will assess their subordinates at lower levels of attribute possession than expected by their subordinates. This may result in loss of calibration in performance evaluations of subordinates, possibly causing conflict in relationships, and possibly affecting turnover intentions. It may even affect audit quality. These possibilities are of importance to accounting firms and are worthy of further investigation in future research.

Limitations

There is no reason to believe that participants had difficulty understanding the definitions of the expert attributes as provided in the appendix. However, one cannot be sure. For example, the attribute “Perceptive” was judged by 40% of participants in [Abdolmohammadi et al. \(2004\)](#) as innate, and 41% as developable. Future research may benefit from more concise definitions and tests of the relationship between attributes and performance measures (e.g., firm performance evaluations). While recent studies (e.g., [McKnight & Wright, 2005](#)) indicate that auditors, particularly experienced ones, have accurate understanding of the determinants of performance, investigation of the relationship between various expert attributes, particularly those not investigated in the past, and superior performance may be a fruitful research direction.

A Likert scale of 1–5 was used to assess attribute possession by auditors. A problem associated with this scale is the possibility of a tendency toward the mean bias. While the results of the study do not indicate this bias, the use of a forced ranking procedure (i.e., select the top five attributes and place them in box (a) and the next five in box (b), etc.) in future research might work more effectively in deciphering differences in attribute possession assessments.

There was also a sample limitation in this study, where 43 subjects were classified by their self-assessed levels of expertise. This procedure resulted in double-digit numbers of subjects classified as specialty candidates ($n = 13$), and competent specialists ($n = 18$). However, only eight subjects were identified as experts. Since, some mixed results were observed by expertise level in the self-peer assessments, future research may be needed to further investigate self-peer assessment issues in audit firms.

Finally, the literature indicates that individuals inflate their own ratings when they rate themselves and their peers. The current study finds this bias only for competent specialists. Expert specialists assessed their own possession at levels lower than their peers, and specialty candidates’ perceived possession of attributes by themselves and their peers did not differ. Given the mixed results, replication of the study to investigate this bias further is indicated. For example, the manager/partner participants in this study on average reported to have had only 4.92 years of experience. Those managers/partners who rated themselves as expert specialists also on average had only 5.38 years of experience. With these relatively low experience levels, these subjects may have compared themselves to “top” experts in the industry and may have underassessed their own possession of

expert attributes. It follows that future studies might benefit from recruiting specialists at higher levels of expertise than those in the current study.

NOTES

1. Extensive reviews of this literature are provided by Bedard (1989), Choo (1989), Colbert (1989), Davis and Solomon (1989), Bonner and Pennington (1991), Bedard and Chi (1993), and Libby (1995).

2. Abdolmohammadi et al. (2004) found the 11th attribute "Quick Thinker" (mean = 3.45) to be significantly more important than the next attribute, "Inquisitive" (mean = 3.15). "Inquisitive" was one of five attributes that the authors included as distracter/filler items (the other four being assertive, energetic, methodical, and perfectionist), all of which were rated below the top 11 attributes (i.e., they were rated only to be moderately, mildly, or minimally important to auditing expertise).

3. These data were not reported in Abdolmohammadi et al.'s (2004) paper. I wish to thank the authors for their permission to use these data in the current study.

4. Superior performance is defined in terms of the efficiency and effectiveness with which auditors perform their tasks (cf., Bedard & Chi, 1993; Davis & Solomon, 1989; Libby, 1995) and studies have reported the effects of some expertise attributes such as ability, knowledge, and experience to investigate superior performance in auditing. Examples of such studies include Bonner and Lewis (1990), Libby and Luft (1993), and Libby and Tan (1994). In practice, superior performance in a rank is generally taken into consideration for promotion of an auditor to higher ranks.

5. Confounding variables such as self esteem (Baird, 1977) and lack of anonymity (Ghorpade, 2000) reduce the value of self, peer, superior, and subordinate assessments. Researchers have suggested conditions to mitigate the effects of these confounding variables. These conditions can be summarized as (1) similar abilities are assessed, (2) measures of skills are detailed (i.e., not general), (3) same time period is used, (4) individuals have experience in assessment, and (5) there is a criterion group (e.g., co-workers) (Mabe & West, 1982). These conditions are met in the current study (see the Methods section).

6. Libby (1995) argues that auditing has distinctive features that distinguish it from other experimental setting typically studied in other fields and there is evidence in support of this argument. For example, in their recent study, Tan & Jamal (2006) report higher meta perception among auditors than those in casual settings in psychology studies.

7. Accounting firms use different terminology for classification of expertise in specialty. The generic classification terminology in this study was meant to elicit information for three levels of expertise in specialty from different firms.

8. A battery of parametric two-sample *t*-tests or analyses of variance, as appropriate, was also performed. The results were generally consistent with those of the non-parametric tests. Also, a battery of Kruskal-Wallis tests of the differences between attribute assessments by levels of specialty at the individual attribute level was performed. This analysis resulted in an interesting pattern. First, there was a high level of agreement about the degree of possession of attributes by expert

specialists. Only two tests (for current knowledge and communicates expertise) indicated differences but only at marginal significance levels. However, there was less agreement with respect to the attributes of lower level specialists, where a systematic bias emerged. Specifically, in comparison to expert specialists, competent specialists consistently underassessed the possession of attributes by specialty candidates and expert specialists. For test of Hypothesis H₃ at the detailed attribute level, a battery of 36 non-parametric Mann–Whitney tests (11 attributes times three specialty levels plus aggregate of all attributes times three specialty levels) comparing the attribute possession assessments of self and peers was performed. While variations were observed between attributes, the overall results were consistent with the aggregate results reported in Table 4.

ACKNOWLEDGMENTS

I gratefully acknowledge participation in this study by professional auditors from several anonymous accounting firms. Thanks to the editor, Vicky Arnold, anonymous associate editor, and two reviewers for their many constructive comments. The chapter has also benefited from comments by Jere Francis, Jim Hunton, Steve Kaplan, Ted Mock, Jerry Searfoss, Jim Shanteau, Jay Thibodeau, Catherine Usoff, Bill Wright, and participants at the 2005 mid-year meeting of the Auditing Section of the American Accounting Association, the 2006 meetings of the Canadian Academic Accounting Association, and the 2006 European Accounting Association. Financial support for this project was provided by Bentley College through summer research support program.

REFERENCES

- Abdolmohammadi, M. J., Searfoss, D. J., & Shanteau, J. (2004). An investigation of the attributes of top industry audit specialists. *Behavioral Research in Accounting*, 16, 1–17.
- Abdolmohammadi, M. J., & Shanteau, J. (1992). Personal attributes of expert auditors. *Organizational Behavior and Human Decision Process*, 53(November), 158–172.
- AICPA. (1988). *Code of professional conduct*. New York: American Institute of Certified Public Accountants.
- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personal Psychology*, 51, 577–598.
- Baird, L. S. (1977). Self and superior ratings of performance; as related to self-esteem and satisfaction with supervision. *Academy of Management Journal*, 20(2), 291–300.

- Bass, B., & Yammarino, F. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review*, *40*, 437–454.
- Bedard, J. (1989). Expertise in auditing: Myth or reality. *Accounting Organizations and Society*, *14*(1/2), 113–131.
- Bedard, J., & Chi, M. T. H. (1993). Expertise in auditing. *Auditing: A Journal of Practice and Theory*, *12*(Suppl.), 21–45.
- Bhamornsiri, S., & Guinn, R. E. (1991). The road to partnership in the “Big Six” firms: Implications for accounting education. *Issues in Accounting Education*, *6*(1), 9–24.
- Bierstaker, J. L., & Wright, S. (2001). A research note concerning problem-solving ability as a predictor of performance in auditing tasks. *Behavioral Research in Accounting*, *13*, 49–62.
- Bonner, S. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society*, *19*(3), 213–234.
- Bonner, S., & Lewis, B. (1990). Determinants of auditor expertise. *Journal of Accounting Research*, *28*(Suppl.), 1–20.
- Bonner, S., & Pennington, N. (1991). Cognitive processes and knowledge as determinants of auditor expertise. *Journal of Accounting Literature*, *10*, 1–50.
- Campbell, D., & Lee, C. (1988). Self appraisal in performance evaluation: Development vs. evaluation. *Academy of Management Review*, *13*, 302–314.
- Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, *36*(3), 459–475.
- Choo, F. (1989). Expert-novice differences in judgment/decision making research. *Journal of Accounting Literature*, *8*, 106–136.
- Colbert, J. L. (1989). The effect of experience on auditors' judgments. *Journal of Accounting Literature*, *8*, 137–149.
- Davis, J. S., & Solomon, I. (1989). Experience, expertise, and expert-performance research in accounting. *Journal of Accounting Literature*, *8*, 150–164.
- Depaulo, B. M., Kenny, D. A., Hoover, C. W., Webb, W., & Oliver, P. V. (1987). Accuracy of person perception: Do people know what kinds of impressions they convey? *Journal of Personality and Social Psychology*, *52*(2), 303–315.
- Emby, C., & Etherington, L. D. (1996). Performance evaluation of auditors: Role of perceptions of superiors and subordinates. *Auditing: A Journal of Practice and Theory*, *15*(2), 99–109.
- Furnham, A., & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology*, *67*, 57–67.
- Ghorpade, J. (2000). Managing five paradoxes of 360-degree feedback. *The Academy of Management Executive*, *14*(February), 140–150.
- Gibbins, M., & Trotman, K. T. (2002). Audit review: Managers' interpersonal expectations and conduct of the review. *Contemporary Accounting Research*, *19*, 411–444.
- Harris, M., & Schaubroeck, V. (1988). A meta analysis of self-supervisor, self-peer and peer-supervisor ratings. *Personal Psychology*, *41*, 43–61.
- Kennedy, J., & Peecher, M. E. (1997). Judging auditors' technical knowledge. *Journal of Accounting Research*, *35*(2), 279–294.
- King, R. (1996). Reputation formation for reliable reporting: An experimental investigation. *The Accounting Review*, *71*(3), 375–395.

- Lawrie, J. W. (1989). Your performance: Appraise it yourself. *Personnel*, 66(1), 21–23.
- Libby, R. (1995). The role of knowledge and memory in audit judgment. In: A. H. Ashton & R. H. Ashton (Eds), *Judgment and decision making research in accounting and auditing* (pp. 176–206). New York: Cambridge University Press.
- Libby, R., & Luft, J. (1993). Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations and Society*, 18(5), 425–450.
- Libby, R., & Tan, H. T. (1994). Modeling determinants of audit expertise. *Accounting, Organizations and Society* (November), 701–716.
- Litras, T. S. (1980). A practical application of peer evaluation. *Research Management*, 23(March), 13.
- Mabe, P., & West, S. (1982). Validity of self-evaluation of ability: A review and meta analysis. *Journal of Applied Psychology*, 67, 280–296.
- Mayhew, B. W. (2001). Auditor reputation building. *Journal of Accounting Research*, 39(3), 599–617.
- McKnight, C. A., & Wright, W. F. (2005). *Different types of knowledge as determinants of the expertise of auditors*. Working paper, University of Waterloo.
- Rich, J. S., Solomon, I., & Trotman, K. T. (1997). The Audit review process: A characterization from the persuasion perspective. *Accounting, Organizations, and Society* (May), 481–505.
- Silverthorne, C. (2000). Situational leadership theory in Taiwan: A different culture perspective. *Leadership & Organization Development Journal*, 21(1/2), 68–75.
- Stone, D. N., Hunton, J. E., & Wier, B. (2000). Succeeding in managerial accounting. Part 1: knowledge, ability, and rank. *Accounting, Organizations and Society*, 25, 697–715.
- Tan, H. T. (1999). Organizational levels and perceived importance of attributes for superior audit performance. *Abacus*, 35(1), 77–90.
- Tan, H. T., & Jamal, K. (2006). Managing perceptions of technical competence: How well do auditors know how others view them? *Contemporary Accounting Research*, 23(3), 761–787.
- Tan, H. T., & Libby, R. (1997). Tacit managerial versus technical knowledge as determinants of audit expertise in the field. *Journal of Accounting Research*, 35(Spring), 97–113.
- Wright, W. F. (2001). Task experience as predictor of superior loan loss judgments. *Auditing: A Journal of Practice and Theory*, 20(1), 147–155.
- Yammarino, F. J., & Atwater, L. E. (1993). Understanding self-perception accuracy: Implications for human resource management. *Human Resource Management*, 32(2/3), 231–247.

APPENDIX. EXPERTISE ATTRIBUTES AND THEIR RELATIVE IMPORTANCE RATINGS – SCALE 1–5 (MINIMALLY IMPORTANT TO EXTREMELY IMPORTANT)

Attribute	Description	Mean (SD)	Trainable/Developable/Innate				
			<i>T</i>	<i>D</i>	<i>I</i>	Designated	
1	Current knowledge	Has an extensive knowledge base. Makes a special effort to keep up with facts, trends, and developments.	4.47 (0.88)	94%	71%	2%	<i>T</i>
2	Problem solver	Is capable of generating new approaches to solving difficult problems. When faced with a new problem, he/she can develop new strategies to solve that problem.	4.16 (0.96)	33%	76%	14%	<i>D</i>
3	Experience	Effectively uses direct and indirect experience to make decisions. Is skillful in making decisions based on past experience.	4.14 (0.99)	36%	91%	3%	<i>D</i>
4	Perceptive	Is able to extract information from a problem that others cannot see. Is insightful in recognition and evaluation of a confusing situation.	4.12 (0.93)	7%	41%	40%	<i>D</i>
5	Communicates expertise	Convinces others that he/she has specialized knowledge. Effectively communicates his/her ability to make decisions to others.	4.10 (1.38)	85%	83%	1%	<i>T</i>

APPENDIX. (Continued)

Attribute	Description	Mean (SD)	Trainable/Developable/Innate				
			<i>T</i>	<i>D</i>	<i>I</i>	Designated	
6	Self-confidence	Has strong belief in his/her ability to make good decisions. Is calm and self-assured while making decisions.	4.01 (1.04)	17%	81%	14%	<i>D</i>
7	Adaptability	Adjusts decision-making strategy to fit current situation. Is responsive to changes in conditions of the on-going problem situation.	3.81 (1.09)	34%	83%	11%	<i>D</i>
8	Intelligence	Has a high level of intelligence. Understands complex problem situations quickly.	3.71 (1.05)	8%	20%	59%	<i>I</i>
9	Knows what is relevant	Readily distinguishes relevant from irrelevant information in a problem. Utilizes only what is relevant; ignores what is irrelevant.	3.65 (1.02)	40%	87%	4%	<i>D</i>
10	Assumes responsibility	Accepts responsibility for the outcomes of decisions, successful or unsuccessful. Is willing to stand behind his/her decisions.	3.63 (1.08)	22%	70%	21%	<i>D</i>
11	Quick thinker	Quickly perceives data relationships. Is able to rapidly envision future possibilities and pitfalls.	3.45 (1.18)	9%	25%	49%	<i>I</i>

Source: Adapted from Abdolmohammadi et al. (2004).

Note: *T*, Trainable; *D*, Developable; *I*, Neither trainable nor developable, thus innate.

DO AUDIT AND NON-AUDIT BUSINESS STUDENTS IMPLICITLY ASSOCIATE A COMPANY'S RELATIVE STOCK MARKET PERFORMANCE WITH PERCEPTIONS OF CORPORATE ETHICAL BEHAVIOR?

Christine Nolder and James E. Hunton

ABSTRACT

Jost et al. (2003) theorizes and finds that business students, on an average, hold a positive fair market ideology (FMI), which suggests that they believe in the power of market forces to reward ethical corporate behavior and punish unethical behavior; accordingly, they tend to make an implicit association between a company's financial performance relative to the stock market and the company's ethics. We suggest that audit education in professional skepticism and 'red flag' analysis will mitigate this implicit bias when a company's relative market performance is unusually distant from a referent benchmark, such as an industry average. In a between-participants experiment involving 94 non-audit and 94 audit business students, we measure their FMI, and examine how they perceive

Advances in Accounting Behavioral Research, Volume 11, 103–128

Copyright © 2008 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11005-5

the ethicality of a company's management based on the referent direction (above or below the industry average) and referent magnitude (relatively close to or distant from the industry average) of the company's relative market performance. The results suggest that both non-audit and audit students indeed hold a positive FMI, and they ascribe favorable ethical perceptions to company performance that is relatively close to the industry average, irrespective of referent direction. When company performance is relatively distant from the industry average, neither group of students makes the implicit link. Overall, the findings do not indicate that audit education differentially affects business students' perceptions of corporate ethics when a company's relative stock market performance deviates considerably from a referent benchmark.

INTRODUCTION

The purpose of the current study is to examine the extent to which audit and non-audit business students ascribe an implicit link between a company's relative market performance and their perceptions of corporate ethical behavior. Jost, Blount, Pfeffer, and Hunyady (2003) theorized and found that business students, in general, implicitly hold a positive ethical bias toward a company that is outperforming a market average. However, audit students are taught to view deviations from a referent benchmark with professional skepticism, particularly if such deviations are deemed to be relatively extreme or unusual. Thus, we hypothesize that audit students will not make an implicit association between relative market performance and perceptions of corporate ethical behavior if the company's performance is relatively distant from a referent benchmark, such as an industry average.

Jost et al. (2003) suggests that favorable beliefs held by business students regarding the efficiency and effectiveness of market forces, termed their *fair market ideology* (FMI), forms the psychological basis for their implicit link between relative corporate performance and perceived corporate ethicality. The FMI bias theorized and tested by Jost et al. (2003) implies an inherent trust in the financial market to act as an ethical regulator; meaning, the market will reward good or ethical companies and punish bad or unethical companies.

The current study posits that audit education will serve to decouple the implicit link between a business student's FMI and their perceptions of corporate ethical behavior when a company's market performance is unusual or unexpected, relative to a referent benchmark. Audit education

stresses the importance of recognizing abnormalities or ‘red flags,’ which include, but are not limited to, management’s attitude, weak internal controls, and unusually rapid growth or profitability relative to the industry (Smith, Omar, Idris, & Baharuddin, 2005; Majid, Gul, & Tsui, 2001). We expect that audit students will view unusually large deviations from an industry average as ‘red flags,’ which should raise the implicit link examined in this study to an explicit level of cognitive reasoning, as audit education teaches them to think through causal mechanisms that underlie unusual or unexpected deviations from referent benchmarks. As a result, we suggest that audit students will not implicitly link financial performance with corporate ethics when a company’s relative financial performance is unusual or distant from the industry average.

A total of 188 undergraduate business students (94 auditing and 94 non-auditing) participated in a between-participant experiment that included two randomized factors (*referent direction* – the case company’s financial performance was above or below the industry average; *referent magnitude* – the case company’s financial performance was relatively close or relatively distant from the industry average), and one measured variable (audit or non-audit students). Given only a 5-year average price/earnings (P/E) ratio for both the case company and the industry as a referent benchmark, participants recorded their perceptions of the company’s corporate ethics. The experimental results suggest that non-audit students ascribed positive corporate ethical behavior to a company when its stock market performance was relatively close to the industry average, regardless of referent direction. When market performance was relatively distant from the industry average, non-audit students indicated no implicit association between market performance and perceived corporate ethical behavior whether relative performance was above or below the industry average. Consistent with our predictions, non-audit students’ responses were similar to audit students’ reactions when referent magnitude was relatively close to the industry average. Contrary to our expectations, though, the responses of non-audit students mirrored those of the audit students when referent magnitude was relatively distant from the industry average.

The current study contributes to extant theory in the following ways. First, our results suggest that gender might explain some variation of FMI among business students, as we found that female students do not hold a favorable FMI, whereas male students’ FMI is favorable. Second, the experimental results are different from the findings of Jost et al. (2003), as they found an interaction between referent magnitude and referent direction, where we only found a positive link between perceived ethicality

and corporate performance when the company's deviation magnitude was relatively close to a referent benchmark. The implication of our finding is that both non-audit and audit business students appeared to decouple the implicit association between relative performance and perceived ethicality when the company's performance deviation was unusual, which signals a positive message about the students' critical analysis and reasoning skills. Third, we used a P/E ratio in the case materials rather than an 'earnings per share' (EPS) ratio employed by Jost et al. (2003). We believe a P/E ratio is more reflective of market performance and the underlying theory; hence, it should be used by future researchers in this area. The next section reviews relevant background literature and theory, and presents the study hypotheses. The following sections present the research method and statistical results. The final section discusses the study findings and offers future research suggestions.

THEORY AND HYPOTHESES

For over 40 years, economic literature has debated the efficient market hypothesis, which generally holds that the financial markets quickly impound all publicly available information into stock prices (Fama, 1970, 1991; Malkiel, 2003). Langevoort (2004) indicates that board members rely on the efficiency of market-determined stock prices, among other indicators, to corroborate the credibility of management's financial reports. Similarly, Jost et al. (2003) suggests that business students tend to believe in the power of market efficiency to punish bad and reward good corporate behavior. However, we caution that over-reliance on market efficiency as a regulator of ethical corporate behavior can engender self-deceptive faith in the integrity of management.

In the current study, we examine the extent to which audit students implicitly link relative stock market performance to perceived ethicality of corporate management. If they exhibit an implicit association of this nature, audit educators and regulators should be concerned about the influence of this implicit bias on audit professionals' fraud awareness, particularly when the stock market performance of a client company is exceeding a referent benchmark, such as an industry average.

Consistent with Jost et al. (2003), related studies suggests that most citizens of a free-market society believe the outcomes of a capitalistic economic system are fair, ethical, and legitimate (Tyler & Lind, 2002; Shiller, 2005). Jost et al. (2003, p. 65) termed this phenomenon FMI.

After developing a psychometric test to measure FMI, Jost et al. (2003) surveyed five different samples of MBA students from the University of Chicago, Stanford University, and New York University, and two non-business student samples from Boston University and Stanford University. Their FMI results indicated that only business students displayed a propensity to believe that market driven outcomes are fair, ethical, and legitimate.

Self-selection bias may partially explain business students' favorable FMI; meaning, college students self-select into business majors based on various motivators and effects. According to Reber (1995), motivators induce individuals to act in ways that are consistent with their internal states of desire; for instance, students' self-selection into business majors can be driven by such motivating factors as aspirations to be successful, expectations of realizing good fortune, and desires for economic security. In turn, motivators to major in business can trigger positive affects toward business-oriented economic philosophies, while classroom experiences can further strengthen such affective attitudes. Hence, in large part, the favorable FMI exhibited by business students appears to be endogenously initiated and exogenously reinforced. We propose the following hypothesis, which is expected to replicate Jost et al. (2003) and serve as a foundation for upcoming hypotheses:

H₁. Audit and non-audit business students will share a similarly positive fair market ideology.

Jost et al. (2003) further suggests that individuals who hold a positive FMI will implicitly link relative stock market performance with corporate ethical behavior. To test their proposition, they randomly surveyed 343 MBA students about their ethical perceptions of both named and unnamed companies whose EPS growth out-performed or under-performed the EPS growth of the S&P 500. Their experimental results indicated that whether the companies were unnamed or named, business students perceived out-performing companies as more ethical than under-performing companies.

Implicit associations of this nature can be examined through the lens of motivators and affects, as they help to shape the nature and extent of cognitive processing (Bonner, 2008). Motivated reasoning theory, for instance, suggests that individuals often bias their cognitive processes to support their desired conclusions or outcomes (Kunda, 1990, 1999). A favorable FMI is indicative of an 'affect heuristic' (Slovic, Finucane, Peters, & MacGregor, 2002; Schwarz, 2002) or biased cognitive process, where the desired outcome is consistent with a socio-economic philosophy that a capitalistic market is good for society. Accordingly, business students

likely allocate relatively little cognitive effort toward ‘reasoning through’ the relationship between a company’s market performance and its ethical behavior. If an individual legitimizes the capital market and trusts the power of market forces to reward socially responsible corporate behavior with fair and legitimate outcomes, it becomes relatively effortless from a cognitive processing standpoint to form an implicit association between market performance and ethical behavior. Motivated reasoning theory and heuristic cognitive processing help to explain the seemingly unwarranted link that many business students exhibit between a positive FMI and perceptions that out-performing companies operating in a free-market system must be ethical (Jost et al., 2003). We are positing the same implicit relationship for non-audit business students:

H₂. Non-audit business students will positively associate a company’s relative stock market performance (above or below the industry average) and their perceptions of the company’s ethical behavior.

Finding empirical support for H₂ will help to strengthen the validity of Jost et al. (2003) through replication and serve as a comparison to audit students, as we suggest that business students who are educated in auditing will not make the implicit link between a company’s market performance and ethical corporate behavior if the deviation is unusual or relatively distant from a referent benchmark, as described below.

Professional Skepticism and Red Flags

Audit students are trained to be professionally skeptical when planning and performing an audit (Payne & Ramsay, 2005). Extant literature does not include a universal definition for professional skepticism, however, the IFAC (2007) section 200.16 states ‘... an attitude of professional skepticism means the auditor makes a critical assessment, with a questioning mind, of the validity of audit evidence obtained and is alert to audit evidence that contradicts or brings into question the reliability of documents and responses to inquiries and other information obtained from management and those charged with governance.’ The IFAC guidance goes on to specifically mention the importance of not overlooking unusual circumstances, as they may be an indication of a risk factor. The rest of the passage highlights how important it is not to apply heuristics, assumptions, or generalizations when evaluating unusual or unexpected deviations from a referent benchmark.

Professional skepticism, as described by IFAC, seems to directly address business students' implicit biases by raising their risk awareness to an explicit level. Audit education reinforces the notion of professional skepticism by focusing on related standards of due care, independence, integrity, and objectivity. Taken as a whole, these concepts comprise the foundation of ethical reasoning and fraud awareness (Jones, Massey, & Thorne, 2003).

A related theme throughout audit education involves the idea of 'red flags.' The identification of risk factors, referred to as 'red flags,' has been touted as an effective fraud detection tool (Pincus, 1989); specifically, once 'red flags' are identified, auditors are instructed to be more watchful, suspicious, and sensitive to the possibility of unethical behavior by management (Pincus, 1989). While extant literature is mixed as to what exactly constitutes a 'red flag,' common examples include unexpected analytical deviations from past performance, management's overall attitude, and unusual or rapid growth relative to the industry (De Berry & Merritt, 2006; Smith et al., 2005; Majid et al., 2001).

The concept of 'red flags' incorporates the notion of referent benchmarks, such as industry comparisons. Auditors are trained to reason through causal mechanisms that underlie unusual or unexpected deviations from referent benchmarks and provide supporting evidence for any inferential conclusions they draw; however, they typically do not exert as much effort into investigating relatively small deviations. Hence, when a company's financial performance is relatively close to the industry average, audit students will likely resort to less effortful heuristic cognitive processing as described earlier, thereby making the same implicit link between relative stock market performance and ethical perceptions as non-audit business students, as reflected in H_{3a} :

H_{3a} . Audit business students will positively associate a company's relative stock market performance (above or below the industry average) and their perceptions of the company's ethical behavior when the deviation magnitude is relatively close to the industry average.

However, when a company's performance is relatively distant from the industry average, or unusual, the combination of professional skepticism, which emphasizes non-reliance on assumptions, generalizations, and heuristics, coupled with 'red flag' awareness, should result in a different finding, as posited below:

H_{3b} . Audit business students will not associate a company's relative stock market performance (above or below the industry average) and their perceptions of the company's ethical behavior when the deviation magnitude is relatively distant from the industry average.

Interaction between Referent Direction and Referent Magnitude

Jost et al. (2003) indicated that for *named* companies, when relative performance was above (below) the market, ethical perceptions of corporate behavior were relatively higher (lower) [referent direction main effect], and companies with smaller deviations were rated as more ethical than those with larger deviations [referent magnitude main effect]. Results for *unnamed* companies, however, suggested an interaction between referent direction and referent magnitude, such that students ascribed a larger difference in ethical perceptions between above and below companies when the deviation magnitude was relatively large, as compared to relatively small.

Specifically, Jost et al. (2003) manipulated a 4% deviation in EPS growth as relatively close to a referent benchmark (S&P 500) and a 13% deviation as relatively distant from the benchmark. Students reflected their ethical perceptions on a scale of 1 (not at all ethical) to 9 (extremely ethical). When the deviation was relatively small, students were unsure about the ethics of an under-performing company (mean = 4.95) and were somewhat positive about the ethics of an out-performing company (mean = 5.49), for a difference of 0.54. When the deviation was relatively large, students were less sure about the ethics of an under-performing company (mean = 4.48) but more positive about the ethics of an out-performing company (mean = 5.66), for a difference of 1.18. The difference-in-differences (1.18–0.54) was significant ($p < .01$). Since the case materials used in the current study involve an unnamed (unrecognizable) company, a similar interaction as reported in Jost et al. (2003) is expected with the non-audit business students, as next hypothesized:

H_{4a}. For non-audit business students, the difference in ethical perceptions of a company that is above or below a referent benchmark (industry average) will be significantly greater when deviation magnitude is relatively distant from, as compared to relatively close to, the industry average.

Recall, H_{3a} predicts that audit students are expected to implicitly link relative financial performance and perceived corporate ethics if the company's performance deviation is relatively close to a referent benchmark. Thus, similar differential responses in ethical perceptions from non-audit and audit students are expected when company performance is above and below the industry average, holding referent magnitude constant at relatively close to a referent benchmark. Contrary to H_{4a} above, for audit students, the differential response between above and below the industry

average should become smaller when referent magnitude is relatively distant from the referent benchmark because audit education in professional skepticism and ‘red flags’ will decouple their implicit link between corporate financial performance and perceived corporate ethics. Hence, we anticipate a different form of interaction for audit students, as follows:

H_{4b}. For audit business students, the difference in ethical perceptions of a company that is above or below a referent benchmark (industry average) will be significantly smaller when deviation magnitude is relatively distant from, as compared to relatively close to, the industry average.

See Fig. 1 for an illustration of H_{4a} (panel A) H_{4b} (panel B).

RESEARCH METHOD

To test the study hypotheses, we conducted a $2 \times 2 \times 2$ between-participants experiment. The quasi-experimental design included two randomized factors (*referent direction* – above or below industry average; *referent magnitude* – relatively close to or relatively distant from the industry average), and one measured variable (audit or non-audit students). Both audit and non-audit students were randomly assigned to one of the four treatments. After the participants responded to dependent variable metrics, they completed manipulation check, debriefing, and demographic items.

Independent Variables

Each participant read information about the five-year average P/E ratio of a case relative to the industry average. The company’s P/E ratio was either above or below (referent direction) the industry average by either 4% or 13% (referent magnitude).¹ In addition, participants were provided with a graph that showed the 4% treatments (above or below the industry average, depending on their randomized condition) within one standard deviation of the industry average or the 13% treatments (above or below the industry average, depending on their randomized condition) beyond two standard deviations from the industry average (see the appendix for an example of the experimental materials). The intent of providing the graph in the case materials was to help the participants visualize a relatively *close* referent magnitude (less than one standard deviation) or a relatively *distant* referent magnitude (outside of two standard deviations).

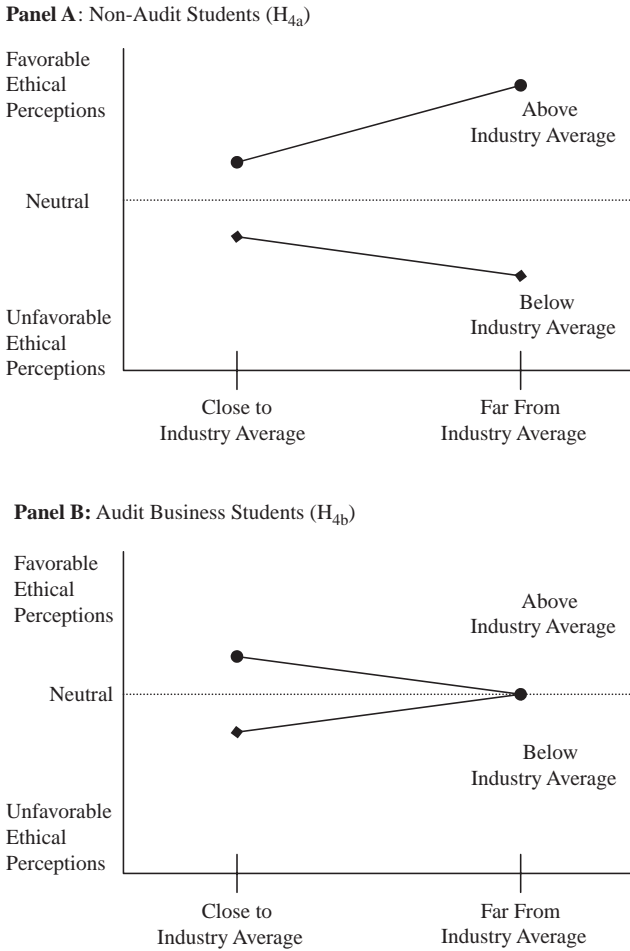


Fig. 1. Illustrations of the Interactive Nature of Hypotheses H_{4a} and H_{4b} .

Dependent Variables

Two dependent variables were assessed in the current study: FMI and *perceived ethicality* of the case company. Regarding FMI, participants responded to a previously tested, shortened version of the FMI psychometric instrument (Jost et al., 2003) that included the following six items: [answers were recorded on an 11-point scale, which was anchored on the

left-hand side by 'strongly disagree' (5), anchored on the right-hand side by 'strongly agree' (5), and denoted as 'not sure' (0) at the midpoint]:²

- (1) The free-market system is a fair system in the long run.³
- (2) Common or 'normal' business practices must be fair, or they would not survive in the long run.
- (3) Acting in response to market forces is not always a fair way to conduct business [*reverse coded*].
- (4) In a free-market system, people tend to get the outcomes that they deserve.
- (5) Profitable businesses tend to be more morally responsible than unprofitable businesses.
- (6) Economic markets do not fairly reward people [*reverse coded*].

The six items were averaged to form a composite FMI index (Cronbach standardized $\alpha = .51$). Using the same reduced version, Jost et al. (2003) attained Cronbach $\alpha = .61$ with one sample ($n = 108$) and $\alpha = .78$ with a second sample ($n = 115$).

Participants also responded to four dependent variable measures reflecting their ethical perceptions of the case company (ABC).⁴ The participants ranked their ethical perceptions on an 11-point scale, where the midpoint of the scale (0) was denoted as 'not sure,' the left-hand side was anchored by 'highly unethical' (5), and the right-hand side was anchored by 'highly ethical' (5). The four response items were as follows:

- (1) How ethical do you believe ABC is in terms of *general business practices*?
- (2) How ethical do you believe ABC is in terms of *fair employee treatment*?
- (3) How ethical do you believe ABC is in terms of *responsibility to consumers*?
- (4) How ethical do you believe ABC is in terms of *environmental considerations*?

The four items were averaged to form a composite 'perceived ethicality' index (Cronbach's α based on standardized items = $.77$).⁵

RESULTS

Participants

A total of 188 students took part in the study. All participants were enrolled in senior level courses offered to accounting and non-accounting majors.

The courses included four audit sections (comprising the 94 audit students), four marketing sections, and one management section (comprising the 94 non-audit students). We collected information confirming that the marketing and management seniors had taken no previous audit courses. The mean (standard deviation) age of all participants was 22 (3.38) years, with a range of 20–46 years. There were 104 male (55%) and 84 female (45%) participants in the sample.

Neither age ($t = 1.25$, $p = .21$) nor gender ($X^2 = 0.09$, $p = .77$) were significantly different between audit and non-audit participants. Age ($F = 1.35$, $p = .26$) and gender ($X^2 = 4.77$, $p = .19$) were also non-significant among treatment conditions. Hence, the randomization of treatments to participants was deemed successful.

Manipulation Checks

After responding to the dependent variable measures, students answered manipulation check items. The first question asked students to indicate the direction of their treatment (above or below the industry average).⁶ Seven of the 188 students chose a direction that was inconsistent with their randomized treatment.⁷ We also asked about the deviation magnitude on an 11-point scale that was anchored on the left-hand side as ‘very close to the average’ (5), anchored on the right-hand side as ‘very far from the average’ (5), and with ‘same’ as the midpoint (0).⁸ The 4% (relatively close) mean (standard deviation) was -0.39 (2.68) and the 13% (relatively distant) mean (standard deviation) was 2.94 (2.19). The results were significantly different from each other ($t = -9.456$, $p < 0.01$). The manipulation check results indicate that participants understood the direction and magnitude of the case company’s P/E ratio relative to an industry benchmark as intended by the experimental treatments; thus, the manipulations are deemed successful.

Preliminary Testing

Preliminary ANCOVA analyses were performed using age and gender as potential covariates. For the FMI index, age was non-significant ($F = 0.48$, $p = .83$), but gender was significant ($F = 17.92$, $p < .01$); thus, gender was retained in further analysis of the FMI index. With regard to the ‘perceived ethicality’ index, neither age ($F = 0.98$, $p = .32$) nor gender ($F = 0.66$, $p = .42$) were significant; hence, they were not included in further analyses.

Visual inspection of residuals on normal probability plots suggested that the data were normally distributed for both dependent variable indices (FMI and perceived ethicality). Levene's test for equality of variances reflected no significant differences among the four treatments for either the FMI index ($F = 1.27, p = 0.29$) or the perceived ethicality index ($F = 0.76, p = 0.52$). Based on normality and homogeneity results, we deemed that the use of parametric statistics was appropriate for both dependent variables.

Hypothesis One

Participants' FMI was measured after the experiment, to avoid the possibility of inducing an unintended demand effect. In making this experimental choice, we were cognizant that the treatments could have affected the FMI responses. Because a person's FMI reflects a deeply held set of inter-related beliefs, goals, and values (Jost et al., 2003), it should be relatively resistant to change; hence, the experimental treatments should not affect the participants' FMI.

Descriptive FMI statistics and t -test results are shown on Table 1 and the results of an ANCOVA model are shown on Table 2. Table 1 reveals that both groups of students reflected a positive FMI; specifically, the non-audit students' grand mean of 0.379 is significantly greater than zero ($t = 2.66, p < 0.01$), as is the audit students' grand mean of 0.550 ($t = 4.58, p < 0.01$). Within the audit and non-audit student samples, the treatment means were not significantly different from each other (audit students: $F = 2.047, p = 0.156$; non-audit students: $F = 1.035, p = .312$). Additionally, the FMI grand means between audit and non-audit students were not significantly different ($t = 0.92, p = 0.36$).

As indicated on Table 2, there is a marginally significant three-way interaction among referent direction, referent magnitude, and student classification ($p = .094$). A detailed analysis of the three-way interaction reveals that the independent variables indeed interacted with the FMI responses – an unexpected finding. The marginally significant three-way interaction arises because the sign of the differences in above versus below industry average (referent direction) between the 4% and 13% conditions (referent magnitude) are reversed for the two student samples. Specifically, for non-audit students (Table 1), the difference in above versus below the industry average is -0.582 (0.168–0.750) in the 4% treatment, and $+0.006$ (0.303–0.297) in the 13% treatment; whereas for audit students, the same differences (above and below the industry average) are $+0.315$ (0.741–0.426)

Table 1. Composite FMI Index Means {Standard Deviations} [Sample Sizes] by Treatment Condition^a.

	Non-Audit Students			Audit Students		
	Referent magnitude 4%	Referent magnitude 13%	Average referent magnitude	Referent magnitude 4%	Referent magnitude 13%	Average referent magnitude
Referent direction – above	.168 {.29} [22] $t=0.57$ $p=0.28$.303 {.28} [24] $t=1.06$ $p=0.14$.236 {.21} [46] $t=1.15$ $p=0.12$.741 {.25} [22] $t=2.95$ $p=0.01$.327 {.24} [24] $t=1.37$ $p=0.09$.534 {.17} [46] $t=3.08$ $p=0.01$
Referent direction – below	.750 {.27} [26] $t=2.77$ $p=.005$.297 {.30} [22] $t=0.99$ $p=0.16$.523 {.20} [48] $t=2.60$ $p=0.00$.426 {.23} [25] $t=1.82$ $p=0.04$.704 {.24} [23] $t=2.89$ $p=0.00$.565 {.17} [48] $t=3.34$ $p=0.01$
Average referent direction	.459 {.20} [48] $t=2.30$ $p=0.01$.300 {.20} [46] $t=1.47$ $p=0.07$.379 {.14} [94] $t=2.66$ $p=0.01$.583 {.17} [47] $t=3.42$ $p=0.01$.516 {.17} [47] $t=3.03$ $p=0.01$.550 {.12} [94] $t=4.58$ $p=0.01$

^aOne-way *t*-tests to determine whether the means are significantly greater than zero.

Table 2. ANCOVA Model for FMI Index (All Participants).

Source	SS	df	MS	<i>F</i>	<i>p</i>
Corrected model	28.994	8	3.624	2.236	.027
Intercept	57.181	1	57.181	35.276	.000
Gender	18.054	1	18.054	11.138	.001
Audit vs. non-audit (AN) ^a	1.579	1	1.579	.974	.325
Referent magnitude (RM) ^b	.587	1	.587	.362	.548
Referent direction (RD) ^c	1.216	1	1.216	.750	.388
AN * RM	.077	1	.077	.047	.828
AN * RD	1.047	1	1.047	.646	.423
RD * RM	.095	1	.095	.059	.809
AN * RM * RD	4.583	1	4.583	2.827	.094
Error	290.148	179	1.621		
Total	360.333	188			
Corrected total	319.142	187			

^aStudent Classification (audit and non-audit).

^bReferent Magnitude describes the relative distance from the industry average.

^cReferent Direction describes above or below performance relative to the industry average.

in the 4% treatment and -0.377 ($0.327-0.704$) in the 13% treatment. Given the width of the 11-point scale used to measure the FMI items (-5 through $+5$) compared to the relatively small absolute differences in the mean responses, any inferences drawn from the marginal three-way interaction would be tenuous.

Interestingly, the ANCOVA model shown in Table 2 indicates that gender significantly influenced the FMI results. To better understand the influence of gender, additional analysis was conducted. The non-audit and audit student samples were split between male and female students, and the results show that only male students endorsed a favorable FMI. Specifically, female FMI means were 0.1047 ($n = 43$) for audit and 0.1098 ($n = 41$) for non-audit students, but the means were not significantly different from zero (female audit students, $t = 0.66$, $p = .25$; female non-audit students, $t = .59$, $p = .27$). Conversely, male audit students average FMI was 0.9118 ($n = 51$) and male non-audit students average FMI was 0.6132 ($n = 53$), and both were significantly different from zero (male audit students, $t = 5.20$, $p < .01$; male non-audit students, $t = 2.96$, $p < .01$). However, the gender effect found with the FMI index did not influence the upcoming hypotheses, as gender was not a significant covariate when the dependent variable reflected the 'perceived ethicality' index. Overall, after considering the marginal three-way interaction and gender effect, H_1 was ostensibly supported, as the grand means for non-audit and audit business students were significantly greater than zero.

Descriptive statistics for the composite perceived ethicality index are shown in Table 3. Using the participants' 'perceived ethicality' index as the dependent variable, a $2 \times 2 \times 2$ ANOVA model (Table 4) indicated a significant main effect for referent magnitude ($F = 15.095$, $p < .01$) and a marginally significant two-way interaction between referent magnitude (close or distant from a referent benchmark) and student type (audit and non-audit) ($F = 3.053$, $p = .08$). All remaining main effects, two-way and the three-way interactions were non-significant ($p \geq 0.19$). Due to the marginally significant two-way interaction between referent magnitude and student type, we split the data between audit and non-audit students, where each sub-sample was subsequently analyzed using 2 (referent direction) by 2 (referent magnitude) ANOVA models.

Hypothesis Two

H_2 suggests that non-audit students will perceive an out-performing company (above a referent benchmark) to be more ethical than an

Table 3. Composite Perceived Ethicality Index Means {Standard Deviations} [Sample Sizes] by Treatment Condition^a.

	Non-Audit Students			Audit Students		
	Referent magnitude 4%	Referent magnitude 13%	Average referent magnitude	Referent magnitude 4%	Referent magnitude 13%	Average referent magnitude
Referent direction – above	.3864 {1.08} [22] $t = 1.46$ $p = 0.07$.1875 {1.42} [24] $t = 0.74$ $p = 0.23$.2826 {1.26} [46] $t = 1.55$ $p = 0.06$	1.1818 {1.33} [22] $t = 4.54$ $p = 0.01$	–.1146 {1.39} [24] $t = –0.46$ $p = 0.32$.5054 {1.50} [46] $t = 2.80$ $p = 0.01$
Referent direction – below	.6923 {1.15} [26] $t = 2.86$ $p = 0.01$.1250 {1.25} [22] $t = 0.47$ $p = 0.31$.4323 {1.22} [48] $t = 2.41$ $p = 0.01$.5700 {1.21} [25] $t = 2.33$ $p = 0.01$	–.1522 {.87} [23] $t = –0.59$ $p = 0.27$.2240 {1.11} [48] $t = 1.27$ $p = 0.10$
Average referent direction	.5521 {1.12} [48] $t = 3.08$ $p = 0.01$.1576 {1.32} [46] $t = 0.86$ $p = 0.19$.3590 {1.23} [94] $t = 2.82$ $p = 0.01$.8564 {1.29} [47] $t = 4.81$ $p = 0.01$	–.1330 {1.16} [47] $t = –0.74$ $p = 0.22$.3617 {1.32} [94] $t = 2.87$ $p = 0.01$

^aOne-way *t*-tests to determine whether the means are significantly greater than zero if positive or significantly less than zero if negative.

under-performing company (below a referent benchmark). As indicated in Table 5, referent direction was not significant for non-audit students ($p = .634$). While the mean of 0.2826 (when P/E ratio was above the industry average; see Table 3) is significantly greater than zero ($t = 1.55, p = 0.06$), the mean of 0.4323 (when the P/E ratio was below the industry average; see Table 3) is also significantly different from zero ($t = 2.41, p = 0.01$). Therefore, in both the above and below treatments, non-audit students ascribed positive ethical perceptions of the case company that were not significantly different from each other. Accordingly, H_2 was not supported.

Further analysis indicates, when the deviation magnitude was relatively close to the industry average, the non-audit participants in the current study recorded significant positive ethical perceptions for both above (mean = 0.3864; see Table 3) and below (mean = 0.6923; see Table 3) performing companies (the means are not significantly different: $t = –0.942, p = .351$). When deviation magnitude was relatively distant (13%) from the industry average, though, the perceived ethicality means were not significantly different from zero in either the above treatment (mean = 0.1875;

Table 4. ANCOVA Model for Composite Perceived Ethicality Index (All Participants).

Source	SS	df	MS	F	p
Corrected model	32.215	7	4.602	3.060	.005
Intercept	24.215	1	24.215	16.103	.000
Audit vs. non-audit (AN) ^a	0.026	1	0.026	.017	.896
Referent magnitude (RM) ^b	22.699	1	22.699	15.095	.000
Referent direction (RD) ^c	0.482	1	0.482	.321	.572
AN * RM	4.591	1	4.591	3.053	.082
AN * RD	2.333	1	2.333	1.522	.215
RD * RM	.124	1	.124	.082	.774
AN * RM * RD	2.601	1	2.601	1.730	.190
Error	270.682	180	1.504		
Total	327.313	188			
Corrected total	302.897	187			

^aStudent classification (audit and non-audit).

^bReferent Magnitude describes the relative distance from the industry average.

^cReferent Direction describes above or below performance relative to the industry average.

Table 5. Non-Audit Student ANOVA Model for Composite Perceived Ethicality Index.

Source	SS	df	MS	F	p	Hypothesis
Corrected model	4.815	3	1.605	1.056	.372	
Intercept	11.315	1	11.315	7.447	.008	
Referent magnitude (RM) ^a	3.432	1	3.432	2.259	.136	
Referent direction (RD) ^b	.347	1	.347	.228	.634	H ₂
RD * RM	.794	1	.794	.522	.472	H _{4a}
Error	136.754	90	1.519			
Total	153.688	94				
Corrected total	141.570	93				

^aReferent Magnitude describes the relative distance from the industry average.

^bReferent Direction describes above or below performance relative to the industry average.

see Table 3) or below treatment (mean = 0.1250; see Table 3) ($t = 0.158$, $p = .87$). In summary, non-audit students were unaffected by referent direction, however, referent magnitude signaled positive ethical cues only when the deviation was relatively close to the industry average.

Hypothesis Three

H_{3a} (H_{3b}) suggests that audit students will [will not] perceive an out-performing (above the industry average) company as more ethical than an under-performing (below the industry average) company when the deviation magnitude is relatively close to [distant from] the industry average. Table 6 indicates that referent magnitude was significant ($p < .01$) for audit students. Further analyses indicate that the perceived ethicality mean of 0.8564 (see Table 3) is significantly greater than zero when deviation magnitude was 4%, but the mean of -0.1330 (see Table 3) is not significantly different than zero when deviation magnitude was 13%.

As with the non-audit student sample, audit students ascribed positive ethical perceptions to a company whether it out-performed (mean = 1.1818; see Table 3) or under-performed (mean = 0.570; see Table 3) the industry average when the referent deviation was relatively close to the industry benchmark (the means are not significantly different: $t = 1.649$, $p = .106$). H_{3a} is supported in that the direction and significance of the two means are consistent with expectations; however, it is important to note that both means are positive.

When referent magnitude was relatively distant from the referent benchmark, audit students did not link perceived ethicality to performance that was either above (mean = -0.1146 ; see Table 3) or below (mean = -0.1522 ; see Table 3) relative to the industry average. While H_{3b} is supported, the audit students also did not implicitly link perceived ethicality with relative

Table 6. Audit Student ANOVA Model for Composite Perceived Ethicality Index.

Source	SS	df	MS	F	p	Hypothesis
Corrected model	27.400	3	9.133	6.138	.001	
Intercept	12.927	1	12.927	8.687	.004	
Referent magnitude (RM) ^a	23.884	1	23.884	16.050	.000	H _{3a} , H _{3b}
Referent direction (RD) ^b	2.472	1	2.472	1.661	.201	
RD * RM	1.933	1	1.933	1.299	.257	H _{4b}
Error	133.928	90	1.488			
Total	173.625	94				
Corrected total	161.327	93				

^aReferent Magnitude describes the relative distance from the industry average.

^bReferent Direction describes above or below performance relative to the industry average.

performance when referent magnitude was 13%, thereby suggesting that audit education had no differential effect in this regard.

Hypothesis Four

H_{4a} suggests that non-audit students' mean responses will reflect an interaction between the two treatments, such that the difference-in-differences in ethical perceptions between an out-performing (above) and under-performing (below) company will be significantly greater when the deviation magnitude is relatively distant from, as compared to relatively close to, the industry average. Table 5 indicates a non-significant interaction between referent direction and referent magnitude for non-audit students ($p = .472$); therefore, H_{4a} was not supported.

H_{4b} suggests that audit students will exhibit a significant interaction between referent direction and referent magnitude, such that the difference between the out-performing (above industry average) and under-performing (below industry average) means will be significantly smaller when the deviation magnitude is relatively distant from, as compared to relatively close to, the industry average. However, Table 6 indicates a non-significant interaction between referent direction and referent magnitude for audit students ($p = .257$); hence, H_{4b} is not supported.

Path Analysis

We tested the relationship between FMI and perceived ethicality, and the relationships among referent direction, referent magnitude, and perceived ethicality, as theorized by Jost et al. (2003). The implication of Jost et al. (2003) is that FMI and perceived ethicality are implicitly positively associated, which suggests a direct path between these two measured variables. Additionally, referent direction and referent magnitude are expected to moderate perceived ethicality, which indicates direct paths from the two independent variables to perceived ethicality. The results are shown on Fig. 2.

The overall statistics suggest a very good fitting model, as CFI and AGFI are both above the recommended benchmark of .90, and RMSEA is below the suggested level of .05 (Kline, 2005). Only two of the three standardized paths' coefficients are significant. The positive sign on the significant path from FMI to perceived ethicality (+.306) suggests that higher FMI scores

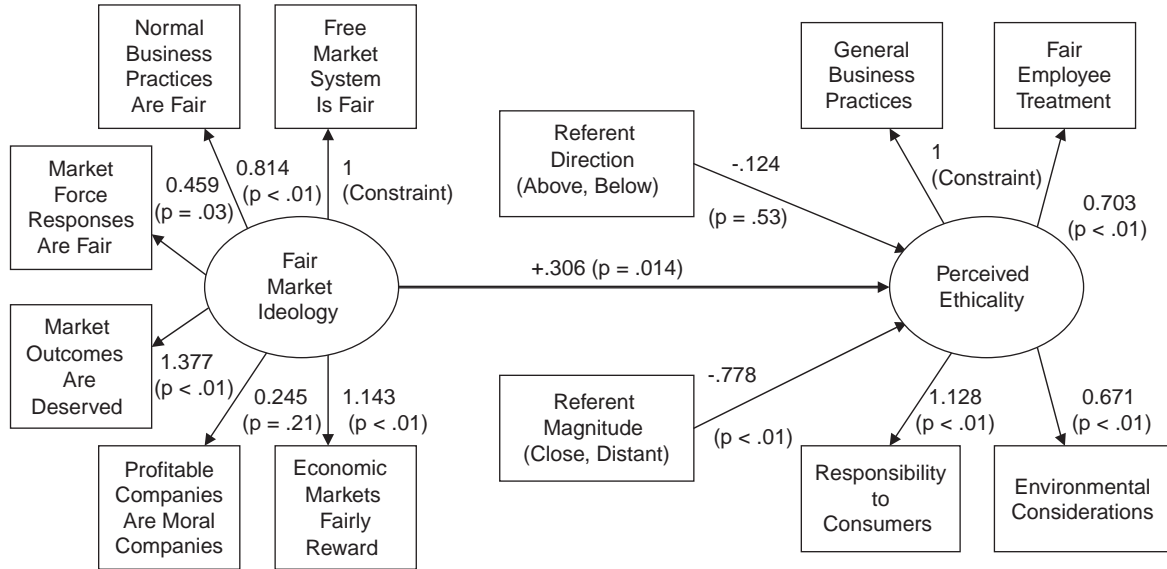


Fig. 2. Path Analysis. Note: Standardized Coefficients are Shown on Paths Leading to Perceived Ethicality. Overall Model Fit Statistics: CFI = .986, AGFI = .929, RMSEA = .020.

were positively associated with higher perceived ethicality scores. The negative path coefficient from referent magnitude to perceived ethicality ($-.778$) indicates that the average positive perceived ethicality score when the referent benchmark was relatively close to the industry average (mean = .459 for non-audit students (Table 1); mean = .583 for audit students (Table 1)) decreased when the referent benchmark was relatively distant from the industry average (mean = .300 for non-audit students (Table 1); mean = .516 for audit students (Table 1)). The non-significant path coefficient from referent direction to perceived ethicality is consistent with the non-significant 'referent direction' main effects reported in the ANOVA models ($p = .634$ for non-audit students (Table 5); $p = .201$ (Table 6) for audit students). Overall, results of the path analysis are consistent with the ANOVA findings reported earlier.

Since our experimental data suggested that the experimental treatments affected FMI and that FMI was associated with gender, we ran some sensitivity analyses to find the best fitting path model. We ran three additional models using the path analysis shown in Fig. 2 as the baseline model: (1) paths from the treatments to FMI were added, (2) a path from gender to FMI was added, and (3) paths from the treatments to FMI and from gender to FMI were added. In each of the three additional models, the overall goodness of fit statistics decreased, relative to the baseline model. Thus, we determined that the baseline model (Fig. 2) is the best fitting model.

Post-Experimental Debriefing

The notion of 'red flags' suggests that audit education should raise an auditor's level of suspicion when he or she identifies an unusual or unexpected deviation from a referent benchmark. To test this relationship, all participants responded to post-experiment debriefing items regarding their level of suspicion about a company's ethical behavior when stock market deviations are slightly lower, slightly higher, unusually lower, and unusually higher than the industry average. For example, the debriefing item regarding an unusually lower deviation read as follows: 'I believe that an unusually lower than average P/E ratio raises suspicion about the ethical behavior of a company.'⁹ Test results were quantitatively similar between the slightly lower and slightly higher items, and between the unusually lower and higher items; thus, they were combined for analysis purposes.

After combining the slightly higher and slightly lower items, results indicated that neither non-audit (mean = -1.84 , SD = 1.95) nor audit

(mean = -1.48 , $SD = 1.92$) business students were suspicious about a corporation's ethical behavior, as both means were significantly less than zero (non-audit students: $t = -9.14$, $p < .01$; audit students: $t = -7.50$, $p < .01$), and the means were not significantly different from each other ($t = 1.26$, $p = .21$). After combining the unusually higher and lower items, audit students' level of suspicion (mean = 0.67 , $SD = 1.83$) was significantly greater than zero ($t = 3.56$, $p < .01$), whereas the level of suspicion recorded by non-audit business students (mean = 0.28 , $SD = 1.94$) was not significantly greater than zero ($t = 1.40$, $p = .16$), but the means were not significantly different from each other ($t = 1.41$, $p = .16$). Hence, we basically find that non-audit and audit business students responded with similar levels of suspicion when a company's stock market performance was relatively close to or distant from a referent benchmark, suggesting that audit education did not result in noticeably different suspicion levels in this study.

DISCUSSION

The objective of the current study is to examine the extent to which audit education will decouple an implicit link between a company's relative market performance and perceptions of corporate ethical behavior when the company's performance deviation is unusually far from a referent benchmark. The theoretical underpinning starts with measuring the participants' FMI, which reflects their underlying belief in the fairness and legitimacy of the financial markets. The findings indicate that both non-audit and audit students indeed hold a favorable FMI, although only male students account for the overall FMI favorability in our sample.

According to [Jost et al. \(2003\)](#), individuals who hold such an ideology will implicitly associate relative market performance and perceptions of corporate ethical behavior; that is, when a company's financial performance exceeds a referent benchmark, such as the industry average, individuals will ascribe positive ethical perceptions to the company and vice versa. These findings do not replicate [Jost et al. \(2003\)](#), as referent direction (above or below an industry average) is not significantly associated with business students' perceived corporate ethical behavior. Furthermore, the results indicate a significant magnitude effect whereby deviations closer to the industry average signal more positive ethical behavior than deviations distant from the industry average. [Jost et al. \(2003\)](#) did not find a significant magnitude main effect for unnamed companies.

The findings do not support the hypotheses that business students who are educated in auditing will respond differently than non-audit students to unusual or distant deviations from a referent benchmark. Instead, the findings show that both audit and non-audit students decoupled the implicit association between relative market performance and perceived ethicality when the company's performance deviation was unusual, relative to a referent benchmark. While this finding was not expected, it sheds a positive light on the critical analysis and reasoning skills of the business students in our sample to 'think through' the causal connection between relative market performance and corporate ethics when a company's market performance is unusual or unexpected.

Several limitations to this study should be recognized. First, while both dependent variables (FMI and perceived ethicality) yielded significant results on an 11-point scale, the absolute values of the results compared to zero were quite small; hence, we are unsure of the practical significance of the implicit link between FMI and perceived ethicality in auditing in particular or the financial marketplace in general. Second, student samples were recruited from one university; thus, the results could be affected by specific curricula, cultural issues, and other unique factors. Finally, company and market background information in the case materials were purposefully omitted to test an implicit link void of any confounding influences; however, had the participants been exposed to a rich set of information related to the company, management, and the market, the research findings might have been different.

We believe that we have introduced an improvement into the FMI literature that will benefit future researchers in this area. Specifically, [Jost et al. \(2003\)](#) proposed that if business students endorse a FMI, they are also likely to believe that more profitable companies are more ethical companies. Their theory was tested using a company's EPS growth relative to the S&P average growth as the dependent variable. We contend that the implied connection between relative EPS growth and profitability is not necessarily straightforward nor consistent since the determination of EPS involves many assumptions, hence it can be manipulated; for instance, variations in EPS across the S&P 500 can be affected by the number of authorized shares by the board of directors, how a company handles the proceeds of treasury stock when dilutive securities are converted into common stock, whether certain contingent convertible bonds are included as outstanding stock, and other assumptions that are used to arrive at diluted EPS. We suggest that future researchers use a company's relative P/E ratio instead. The P/E ratio typically includes a company's past performance (EPS over the last

12 months) as the denominator and the company's current stock price as the numerator. Since the P/E ratio impounds earnings history with future expectations (market value per share), comparison of the P/E ratio to a referent benchmark (e.g., the industry average P/E ratio) signals the extent to which a company is under- or over-performing relative to the market. Therefore, testing associations among FMI and other factors to relative P/E ratios better reflects the underlying theory espoused by Jost et al. (2003).

The research results obtained herein have revealed more suggestions for FMI researchers. Since our findings indicate that gender affected the participants' FMI, subsequent research could examine cognitive, affective, and motivational reasoning behind the female students' relatively neutral FMI and the male students' relatively favorable FMI. Future research might also examine additional factors that could affect FMI, such as socio-economic status, political ideology, age, experience, and profession. Finally, business and audit educators should reinforce to students that a company's increasing profits or improving relative market performance by itself, typically carries no signal, either positive or negative, about the company's ethical behavior.

NOTES

1. Jost et al. (2003) also used 4% and 13% deviations to reflect small (close to) and large (distant from) deviations from a referent benchmark, respectively.

2. Jost et al. (2003) anchored their scales with 'completely disagree' and 'completely agree.'

3. The authors added the phrase 'in the long run' to both questions 1 and 2 as not to confound the instrument with the idiosyncrasies of market behavior in the short term.

4. The ethicality instrument used in the current study was developed by Jost et al. (2003).

5. Jost et al. (2003) did not report a Cronbach alpha statistic for the four item measure of 'perceived ethicality.'

6. The case instructions strongly insisted that the students were not to look back to the preceding case material pages when answering the manipulation check items.

7. Dependent variable results were quantitatively and qualitatively similar whether the 7 were eliminated or included; thus, we retained them in the sample.

8. For this and all subsequent 11-point scales, we coded the left-hand side as negative 1 through 5, and the right-hand side as positive 1 through 5. We did not include negative or positive signs on the scales provided to the participants, as a negative sign can carry a negative connotation.

9. Participants recorded their responses on an 11-point scale, which was anchored on the left-hand side by "strongly agree" (5), anchored on the right-hand side by 'strongly disagree' (5), and denoted as 'not sure' (0) at the midpoint.

REFERENCES

- Bonner, S. E. (2008). *Judgment and decision making in accounting*. New Jersey: Pearson Prentice Hall.
- De Berry, T., & Merritt, M. (2006). A study of the abilities of accounting and auditing professionals in recognizing red flags of fraud. *Journal of Forensic Accounting, 11*, 89–108.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*, 383–417.
- Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance, 46*, 1575–1617.
- International Federation of Accountants (IFAC). (2007). *Handbook of international auditing, assurance and ethics pronouncements*. Available at <http://www.ifac.org/Store>
- Jones, J., Massey, D., & Thorne, L. (2003). Auditors' ethical reasoning: Insights from past research and implications for the future. *Journal of Accounting Literature, 22*, 45–103.
- Jost, J., Blount, S., Pfeffer, J., & Hunyady, G. (2003). Fair market ideology: Its cognitive motivational underpinnings. *Research in Organizational Behavior, 25*, 53–91.
- Kline, R. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge: The MIT Press.
- Langevoort, D. C. (2004). Resetting the corporate thermostat: Lessons from the recent financial scandals about self-deception, deceiving others and the design of internal controls. *Georgetown Law Journal, 93*, 285–317.
- Majid, A., Gul, F., & Tsui, J. (2001). An analysis of Hong Kong auditors' perceptions of the importance of selected red flag factors in risk assessment. *Journal of Business Ethics, 32*(3), 263–274.
- Malkiel, B. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives, 17*(1), 5–82.
- Payne, E., & Ramsay, R. (2005). Fraud risk assessments and auditors' professional skepticism. *Managerial Auditing Journal, 20*(3), 321–330.
- Pincus, K. (1989). The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations and Society, 14*, 153–163.
- Reber, A. (1995). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Schwarz, N. (2002). Feelings as information: Moods influence judgments and processing strategies. In: T. Gilovich, D. Griffin & D. Kahneman (Eds), *Heuristics and biases: The psychology of intuitive judgment* (pp. 534–547). New York: Cambridge University Press.
- Shiller, R. (2005). *Irrational exuberance* (2nd ed.). New Jersey: Princeton University Press.
- Slovic, P., Finucane, M., Peters, E., & MacGregor, D. (2002). The affect heuristic. In: T. Gilovich, D. Griffin & D. Kahneman (Eds), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.
- Smith, M., Omar, N. H., Idris, S. I. Z. S., & Baharuddin, I. (2005). Auditors' perception of fraud risk indicators: Malaysian evidence. *Managerial Auditing Journal, 20*(1), 73–85.
- Tyler, T., & Lind, A. (2002). Understanding the nature of fraternalistic deprivation: Does group-based deprivation involve fair outcomes or fair treatment? In: W. Smith (Ed.), *Relative deprivation: Specification, development and integration* (pp. 44–68). New York: Cambridge University Press.

APPENDIX A. PARTIAL EXPERIMENTAL MATERIALS

Above (Referent Direction) by Close (Referent Magnitude) Condition

Attached please find information on ABC Corporation's average *P/E ratio* compared to the industry average for the past 5 years. Please *carefully* review and *think* about the information, then answer the 4 questions that follow.

The *P/E ratio* of a stock is a measure of the price paid for a share relative to the income or profit earned by the corporation per share. A higher *P/E ratio* means that investors are paying more for each unit of income.

The price per share (numerator) is the market price of a single share of the stock. The EPS (denominator) is the net income of the company for the most recent 12-month period, divided by number of shares outstanding.

The dotted lines on the graph represent 1 and 2 standard deviations from the 5-year industry average *P/E ratio*. For example, if a company's *P/E ratio* is exactly the 5-year industry average, it would fall on the middle line labeled zero.

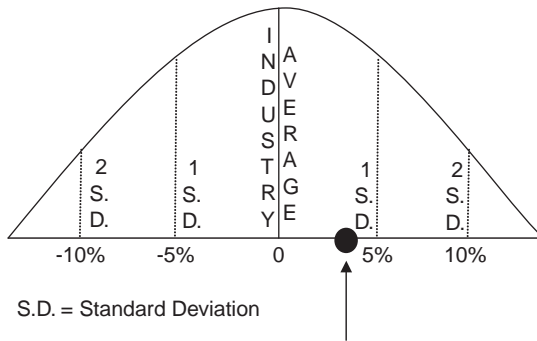


Fig. A1. ABC Company's 5-Year Average *P/E Ratio* is 4% Higher than the Industry's 5-Year Average *P/E Ratio*.

AN EXAMINATION OF THE INFLUENCE OF CONTEXTUAL AND INDIVIDUAL VARIABLES ON PUBLIC ACCOUNTANTS' EXHAUSTION

Daniel W. Law, John T. Sweeney and
Scott L. Summers

ABSTRACT

Despite its widespread acceptance and application in the psychology literature, exhaustion, the core dimension of job burnout, has only recently been examined in the domain of public accounting. These studies highlighted the problem of exhaustion within the profession and examined its causes relative to the environment of public accounting. Another factor, not previously addressed in the context of public accounting, is the role personality plays on public accountants' exhaustion. The current study addressed this void by examining how the personality traits of hardiness, workaholism, neuroticism, and Type-A behavior in public accountants affect exhaustion. The results indicated that public accountants who were high in hardiness experienced significantly less exhaustion. The role stressors of overload and conflict were also significant contributors to public accountants' exhaustion.

Advances in Accounting Behavioral Research, Volume 11, 129–153

Copyright © 2008 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11006-7

INTRODUCTION

The public accounting workplace has long been acknowledged as a high-stress environment (Weick, 1983), motivating a substantive body of stress-related research. The accumulated findings of this research stream generally indicated that high levels of stress existed in public accounting, often resulting in negative consequences to both the firm and to the individual (Collins & Killough, 1992; Haskins, Baglioni, & Cooper, 1990; Smith, Everly, & Johns, 1993). More recently, accounting researchers have concentrated on a specific form of stress known as “burnout,” a condition marked by prolonged negative responses to chronic stressors on the job (Maslach, Schaufeli, & Leiter, 2001; Maslach & Jackson, 1986). Burnout is widely accepted in the psychology, sociology, and organizational behavior literatures as a critical workplace stress syndrome (Cordes & Dougherty, 1993; Halbesleben & Buckley, 2004).

Recently, researchers have focused on the relationship between employee burnout and working conditions in public accounting firms (Almer & Kaplan, 2002; Fogarty, Singh, Rhoads, & Moore, 2000; Sweeney & Summers, 2002). The current study extends this research stream by examining the influence of personality on exhaustion,¹ the central dimension of the burnout syndrome. Exhaustion is best understood as a condition in which job demands produce in the worker a state of depleted emotional and mental resources (Maslach et al., 2001; Moore, 2000a; Cordes, Dougherty, & Blum, 1997).

To examine the relationship between exhaustion and personality, a field survey was administered to a sample of 112 public accountants. The path models indicated that public accountants who were relatively higher in the personality trait of hardiness experienced significantly less exhaustion. No significant relationship between exhaustion and the personality traits of workaholism, Type-A behavior, or neuroticism were found. Consistent with prior research (Almer & Kaplan, 2002; Sweeney & Summers, 2002), the results indicate that public accountants have relatively high levels of exhaustion, and that the stressors of role overload and role conflict positively impact exhaustion. These results also support the prior findings of Sweeney and Summers (2002) that the relationship between public accountants’ workload and exhaustion may be different during the non-busy season than during the busy season. Additionally, these findings show that public accountants who are high in hardiness may have a comparative advantage over their low-hardiness colleagues in meeting the challenges of a stressful workplace.

This study contributes to the body of accounting research on several dimensions. First, we extend the accounting stress literature by introducing the

personality traits of workaholism, hardiness, and neuroticism. These traits, along with Type-A personality, were selected for inclusion because prior research in accounting, psychology, and organizational behavior indicated that they would be especially applicable for the study of stress in public accounting. Importantly, the results of this study suggest that public accountants who are high in hardiness may have a comparative advantage over their low-hardiness colleagues in meeting the challenges of a stressful workplace.

Second, prior accounting studies of the relationship between personality and stress have been limited to one or two personality traits (Choo, 1986; Fisher, 2001; Haskins et al., 1990). This approach can be problematic because personality is multifaceted, and some traits with dissimilar labels intercorrelate so highly that they essentially can be considered measures of an equivalent construct (Judge, Erez, Thoresen, & Bono, 2002; Watson & Clark, 1984). By concurrently examining multiple personality measures, redundant and overlapping constructs can be identified as well as traits that contribute uniquely to public accountants' exhaustion.

Third, this study focuses on exhaustion, the core component of the burnout syndrome and increasingly the singular focus of stress research in psychology and occupational behavior (Babakus, Cravens, Johnston, & Moncrief, 1999; Moore, 2000a, 2000b; Houkes, Janssen, de Jonge, & Bakker, 2003). As the trigger and central construct of the burnout process, exhaustion has become synonymous with burnout and the dimension unanimously accepted by researchers (Cordes & Dougherty, 1993; Maslach et al., 2001). In moving the focus of accounting burnout research to exhaustion, and in extending the literature beyond the work environment–exhaustion dynamic, the results of this study may be helpful to management of public accounting firms in designing and implementing strategies intended to reduce exhaustion.

This chapter proceeds in the following manner. First, a review of the relevant literature and the development of the study's hypotheses are presented. A discussion of methodology and results follow. Finally, the chapter concludes with a discussion of the results and the study's implications.

REVIEW OF PRIOR LITERATURE AND HYPOTHESES

Exhaustion: The Stress Dimension of Burnout

Academic interest in the stressful nature of public accounting appeared to peak in the 1980s and early 1990s, as evidenced by the number of

stress-related articles published during this period. Researchers employed a variety of measures and proxies in attempting to assess the stress of public accountants, including the Stress Arousal Scale (Smith et al., 1993), stress perceptions (Snead & Harrell, 1991), and the Job-Related Tension Scale (Choo, 1986). Some studies did not attempt to measure stress, but instead examined the association between stressors, such as role ambiguity and role conflict, and outcomes (Fisher, 2001; Senetra, 1980). Inconsistencies in stress measures across studies and in modeling the stressor–stress dynamic may have hindered the development of a theoretically consistent, coherent, and generalizable stream of research. By the late 1990s, relatively few stress research articles were appearing in the academic accounting literature, despite significant changes in the public accounting environment.

Partially to address these limitations, recent stress studies in accounting (Almer & Kaplan, 2002; Fogarty et al., 2000; Sweeney & Summers, 2002) have focused on “burnout” – a psychological stress syndrome widely represented in the psychology, sociology, and organizational behavior literatures (Cordes & Dougherty, 1993; Maslach & Jackson, 1986; Maslach et al., 2001). Burnout is a protracted negative response to persistent emotional and interpersonal stressors at work (Maslach et al., 2001). Scores of empirical studies have established burnout, and more specifically its critical component of exhaustion, as an important and generally accepted construct in the area of work stress (Cordes & Dougherty, 1993; Maslach et al., 2001).

Burnout has generally been characterized by three sequentially occurring conditions: emotional exhaustion, depersonalization, and reduced personal accomplishment (Cordes et al., 1997; Lee & Ashforth, 1996; Maslach et al., 2001; Maslach & Jackson, 1986). Despite the large volume of burnout research, however, a lack of consensus has existed regarding the sequential development of the burnout syndrome (Cordes et al., 1997; Golembiewski, Boudreau, Munzenrider, & Luo, 1996). Furthermore, some researchers have questioned whether the depersonalization and reduced personal accomplishment constructs represent components of the syndrome or are actually consequences of exhaustion (Bakker, Van Der Zee, Lewig, & Dollard, 2006; Golembiewski et al., 1996; Moore, 2000a, 2000b; Shirom, 1989; Leiter & Durup, 1996; Koeske & Koeske, 1989). Widespread agreement does exist, however, regarding the centrality of the exhaustion construct to burnout (Maslach et al., 2001; Moore, 2000a) and increasingly stress researchers have focused exclusively on exhaustion (Babakus et al., 1999; Moore, 2000a, 2000b; Houkes et al., 2003; Witt, Andrews, & Carlson, 2004; Wright & Cropanzano, 1998).

Exhaustion, defined as the “depletion of emotional and mental energy needed to meet job demands” (Moore, 2000a, p. 336), represents the basic stress dimension and is the most thoroughly analyzed and reported component of burnout (Maslach et al., 2001). A worker who describes herself as experiencing burnout is generally referring to feelings of exhaustion (Maslach et al., 2001). The exhausted worker feels overextended, experiences a diminution of energy, and often abhors the thought of going to work (Cordes & Dougherty, 1993; Gaines & Jermier, 1983; Lee & Ashforth, 1993; Maslach, 1982; Maslach et al., 2001).

Due to its chronic and intensely affective nature (Gaines & Jermier, 1983), exhaustion carries with it only negative outcomes (Cordes & Dougherty, 1993). Workers who experience high levels of exhaustion often find their passion and commitment for work replaced by feelings of frustration and anxiety (Maslach & Leiter, 1997). Exhausted individuals can suffer from physical and mental problems and strained relationships with co-workers and family members (Maslach & Jackson, 1986; Maslach et al., 2001). Organizations are adversely affected by exhausted workers through lower productivity, absenteeism, and higher turnover (Cordes & Dougherty, 1993; Maslach & Leiter, 1997). As exhaustion is central to and synonymous with the burnout syndrome (Maslach et al., 2001), it is the focus of this study.²

Based on prior literature, the current study proposes a research model linking the impact of work environment and personal variables on exhaustion. Each of the components of the model is discussed below.

Exhaustion and the Public Accounting Work Environment

The relationship between job demands and exhaustion has been extensively examined in the disciplines of psychology and organizational behavior, and more recently in the accounting research literature. Fogarty et al. (2000) first introduced “burnout” to the accounting research literature. They found that burnout tendencies were a key mediator of the relationship between role stressors (role conflict, ambiguity, and overload) and the outcome variables of job satisfaction, job performance, and turnover intentions for a sample of public accountants. Almer and Kaplan (2002) examined the effect of flexible work arrangements on public accountants’ burnout by comparing a sample of CPAs under a standard work arrangement to a demographically similar sample under a flexible arrangement. Extending Fogarty et al. (2000), the authors modeled burnout as mediating the relationship between role stressors and job outcomes. Their results indicated that public accountants under a standard

work arrangement reported higher burnout, lower job satisfaction, and higher turnover intentions than their counterparts under a flexible work schedule.

In a longitudinal study, [Sweeney and Summers \(2002\)](#) examined the effect of the incremental busy-season workload on public accountants' burnout. Hours worked and role stressors were measured for a sample of public accountants immediately prior to and at the conclusion of the busy season. During the busy season, subjects worked an average of 63 h per week, versus 49 h during the pre-busy-season period. During both the pre-busy season and the busy season, the stressors of role overload and role conflict were directly related to burnout. Sweeney and Summers found that the additional workload of the busy season caused public accountants' exhaustion to rise to levels rarely reported in the burnout research literature.

Based upon the results of prior research, the following relationships among job demands and exhaustion are posited:

H₁. The stressors of role overload, role conflict, and role ambiguity will be positively related to exhaustion.

H₂. Workload will be positively related to exhaustion.

Exhaustion and Personality

Personality refers to "dispositions to relate with the world and interact with it in particular ways" ([Demetriou, Kyriakides, & Avraamidou, 2003](#)). Personality traits reflect different modes of relating to the situation or the environment ([Demetriou et al., 2003](#)) and generally remain consistent over time ([Maslach et al., 2001](#); [Spector & O'Connell, 1994](#)). An individual's response to external stimuli, such as workload demands and role stressors, may be mediated by aspects of her or his personality ([Halbesleben & Buckley, 2004](#)). Research has indicated that personality traits influence how an individual responds to a stressful workplace ([Bakker et al., 2006](#); [Cano-Garcia, Padilla-Munoz, & Carrasco-Ortiz, 2005](#); [Maslach et al., 2001](#); [Tokar, Fischer, & Subich, 1998](#)). Personality differences may partially explain why some public accountants have difficulty adjusting to the demands of the job while others seem to prosper under conditions normally associated with high burnout. Investigation of the personality–exhaustion relationship in the environment of public accounting is a logical extension of the existing body of accounting burnout research.

Numerous dimensions and types of personality have been characterized in the psychology literatures ([Lampe, 2004](#)). The sheer number of personality traits identified requires researchers to select those traits that are most likely

to influence the attitude, behavior, or syndrome studied. As such, the following discussion presents applicable background and research studies of personality traits that have been associated with job stress and/or exhaustion. Based on a review of the stress literature in accounting, psychology, and organizational behavior, four personality traits were selected that could potentially mediate the effect of workload and job stressors on exhaustion in public accounting: hardiness, workaholism, Type-A, and neuroticism.

Hardiness

Maddi and Kobasa (1984, p. 50) define hardiness as “a general sense that the environment is satisfying, which leads a person to approach life experiences with curiosity and enthusiasm or commitment.” Hardiness represents an integration of three closely interrelated factors – control, commitment, and challenge (Sheard & Golby, 2007). Through cognitive appraisal, a hardy person perceives potentially stressful situations as meaningful and interesting (commitment), perceives stressors as changeable (control), and views change as a normal part of life and an opportunity for personal growth instead of a threat (challenge). These perspectives enable the hardy person to more effectively cope with job-related stressors. In contrast, individuals low in hardiness often view life experiences as boring and meaningless, and endure more negative reactions to stressful events (Kobasa, Maddi, & Courington, 1981). This may explain why Law (2005), in the only previously published accounting study to examine hardiness, found a significant inverse relationship between public accountants’ level of hardiness and intent to turnover.

In the public accounting environment, individuals higher in hardiness may be better able to manage the stressful working conditions and thereby experience less exhaustion. A number of non-accounting studies have linked lower levels of *hardiness* to higher levels of exhaustion (Constantini, Solano, Di Napoli, & Bosco, 1997; de Vries & van Heck, 2000; Martin, Kelley, & Eklund, 1999; McCranie, Lambert, & Lambert, 1987; Pierce & Malloy, 1990; Rich & Rich, 1987). An inverse relationship between the personality trait of hardiness and public accountants’ exhaustion is expected in the current study.

H₃. Public accountants high in hardiness will experience less exhaustion.

Type-A

Individuals with a Type-A personality exhibit behaviors that include the dimensions of competition, a time-pressured lifestyle, hostility, and an

excessive need for control (Maslach et al., 2001). Other indicators of Type-A behavior include explosive, accelerated speech, high achievement, polyphasic lifestyle, impatience with slowness, heightened pace of living, and a general appearance of tension (Jamal, 1990). Many of these indicators have been empirically identified in various occupations (Matthews, 1982; Price, 1982). Type-A behavior has been directly linked to exhaustion (Maslach et al., 2001; Nagy & Davis, 1985) with non-accounting subjects. In the accounting literature, Type-A behavior has been positively associated with role stressors (Fisher, 2001) and job-related tension (Choo, 1986), as well as mediating the effect of workload stress (Haskins et al., 1990). Based on the results of prior studies, Type-A public accountants may be more likely to experience high levels of exhaustion.

H₄. Accountants high in Type-A behavior will experience greater exhaustion.

Neuroticism

Neuroticism represents a tendency for a person to be highly reactive emotionally (Mayes, Johnson, & Sadri, 2000) and is one of the “Big Five” personality factors, generally accepted among psychology researchers as the five major dimensions of personality (Mount & Barrick, 1995; Saucier, 1994).³ Neurotic individuals worry, are quick to anger, are easily embarrassed, and tend to be depressed and dissatisfied overall with life (Eysenck & Eysenck, 1963; Maslach et al., 2001). Neurotic individuals are emotionally unstable and vulnerable to stress, and are therefore likely to experience greater exhaustion (de Vries & van Heck, 2000; Maslach et al., 2001).

Of the Big Five factors, neuroticism is the most widely studied in the psychology literature (Judge et al., 2002) and the only factor that is consistently associated with exhaustion (de Vries & van Heck, 2000). Of the Big Five factors, Bakker et al. (2006) found neuroticism to be the sole predictor of exhaustion for a sample of counselors, and Zellars, Perrewe, and Hochwarter (2000) found identical results with health care workers. Neuroticism has also been found related to exhaustion for samples of nurses (Allen & Mellor, 2002), service case managers (Goodard, Patton, & Creed, 2004), and hotel employees (Kim, Shin, & Umbreit, 2007). Public accountants high in neuroticism are expected to experience greater exhaustion.

H₅. Public accountants high in neuroticism will experience greater exhaustion.

Workaholism

Spence and Robbins (1992, p. 161) define workaholism as an addiction where the individual “feels driven or compelled to work, not because of external demands or pleasure in work, but because of inner pressures that make the person distressed or guilty about not working.” Workaholism is conceptualized as similar to other addictions (e.g., alcoholism) and has been described as a “progressive, fatal disease” (Fassel, 1990, p. 13). Workaholics realize that their work commitment is excessive, but feel unable to control or reduce it. Spence and Robbins (1992) found workaholics to experience greater levels of stress and more health complaints than other workers. Although workaholism may not directly impact exhaustion, public accountants high in workaholism may be more prone to work longer hours, and therefore experience higher levels of exhaustion resulting from a greater workload (Sweeney & Summers, 2002).

H₆. Workaholism moderates the impact of workload on exhaustion.

METHOD

Sample

The sample of public accountants was acquired from two public accounting organizations representing a multi-office national firm and a multi-office regional firm. All of the subjects were based in the Pacific Northwest of the United States, and the project was approved by the relative institutional review board for the study of human subjects.

The study consisted of a between-subjects design utilizing a field survey incorporating established measures of the variables of interest. Representatives from each of the firms requested that the research instruments be mailed in bulk directly to them. Sample selection was not randomized as all firms agreed to a predetermined limit of participants, and the firm representatives distributed the self-report surveys independent of the researcher. However, the representative from the regional firm (Firm A) did indicate that nearly all of the firm’s public accountant employees received a survey. Respondents were assured that their participation was voluntary and that responses would remain anonymous. Pre-stamped, pre-addressed envelopes were included for the convenience of both the participants and the researchers. After receiving the surveys, almost all

respondents completed and mailed them within a short time frame of about two weeks.

One hundred and twelve research instruments were distributed to public accountants of the regional firm (Firm A) approximately one month after the end of the audit and tax busy season. Seventy-five instruments were returned directly to the researchers for a response rate of 67%, which is generally considered an acceptable rate for survey data (Babbie, 1990). The national firm (Firm B) provided subjects for the study during November and December, just prior to the start busy season. As the interval between data collection for Firms A and B was approximately six months, differences between firms may be confounded with the survey time period. Of the 90 instruments distributed to professionals in Firm B, 40 were returned for a response rate of 44%. The overall response rate of 57% is considered adequate (Babbie, 1990), suggesting that non-response bias is not a concern.⁴

Of the 115 research instruments returned, three respondents from Firm A failed to reply to significant blocks of items. These omissions rendered the instruments unusable for hypotheses testing, resulting in a final sample of 112 public accountants (72 and 40 from Firms A and B, respectively). Based on the number of model variables, this sample size was adequate for statistical testing (Nunnally & Bernstein, 1994).

Table 1 presents demographic data for the sample. The average age of the participants was 36, and the time employed with the current firm and total time employed in public accounting was seven and ten years respectively. Fifty-four percent of the respondents were female, and more than 87% of the participants were married. An examination of Table 1 reveals no significant firm differences relative to the two participating firms, with the exception of rank. Although years in public accounting and years with the current firm are not significantly different between the firms, Firm B was weighted more heavily at the senior ranks. All eight participating partners were from Firm B. The primary functional areas represented in the sample included audit, tax, and consulting, but most participants indicated specializing in tax, followed by auditing.

Measures

Exhaustion was measured using the emotional exhaustion subscale of the Maslach Burnout Inventory (MBI) (Maslach & Jackson, 1986; Maslach, Jackson, & Leiter, 1996). The MBI is the most well-accepted and

Table 1. Demographics of Sample Participants.

Variable	Overall		Firm A		Firm B	
	Mean	SD	Mean	SD	Mean	SD
Age	36.15	10.48	36.33	10.15	35.83	11.15
Years with firm	6.92	7.39	6.22	6.68	8.16	8.48
Years in public accounting	9.83	8.93	9.69	8.54	10.08	9.73

Variable	Overall		Firm A		Firm B	
	Male	Female	Male	Female	Male	Female
Gender (%)	46	54	49	51	40	60

Variable	Overall		Firm A		Firm B	
	Married	Single	Married	Single	Married	Single
Marital status (%)	87	13	89	11	85	15

Rank	Staff	Senior	Supervisor	Manager	Senior manager	Partner
Firm A	18	11	13	27	3	0
Firm B	9	6	N/A	13	4	8

Note: Overall, $n = 112$; Firm A, $n = 72$; Firm B, $n = 40$.

validated measure of exhaustion in the literature (Cordes & Dougherty, 1993; Maslach, Jackson, & Leiter, 2001) and is the measure most utilized in recent accounting studies (e.g., Almer & Kaplan, 2002; Sweeney & Summers, 2002). Cronbach’s coefficient alpha has been calculated at 0.90 (Maslach & Jackson, 1986). The scale has also demonstrated test–retest reliability and factorial, convergent, and discriminant validity (Maslach & Jackson, 1981, 1986). Subjects indicate the frequency of experiencing job-related exhaustion on a 7-point Likert scale, with endpoints 1 = “never occurs” to 7 = “every day.”

An instrument developed by Saucier (1994) was used to measure the personality trait of neuroticism. The scale, developed from a version of Goldberg’s (1992) unipolar Big Five markers, consists of eight markers or unipolar adjectives (fretful, jealous, moody, unenvious, relaxed, temperamental, touchy, and envious). Individuals are asked to evaluate themselves relative to these adjectives using a 7-point continuum with endpoints being “extremely accurate” and “extremely inaccurate.” In a number of independent samples, Cronbach’s coefficient alphas were measured in

the .70's and .80's for this abbreviated scale, and items from the scale corresponded closely with Goldberg's larger set. Psychometric properties were consistent among the independent samples.

A scale modified from Kobasa's (1979) pioneering measure was employed to assess hardiness (Bartone, Ursano, Wright, & Ingraham, 1989). All three closely related dispositional tendencies of hardiness (commitment, challenge, and control) were measured using a scale consisting of 30 items. Respondents were asked to describe themselves using a 4-point scale with endpoints of "not at all true" and "completely true." The modified scale demonstrated good internal reliability (Cronbach's coefficient alpha: 0.82) and correlated highly (0.93) with Kobasa's original scale. This measure is considered a third-generation scale with a number of improvements over earlier scales (Funk, 1992).

Bortner's (1969) scale was used to assess Type-A behavior. This scale utilizes extreme adjectives or descriptors (14 pairs) and asks participants to rate themselves on a continuum between the two logical extremes (e.g., never late, casual about appointments). The continuum is divided equally into 10 for scoring purposes. The scale has demonstrated a moderate degree of inter-item reliability (0.68; Bortner, 1969).

A measure of workaholism entitled "the work addiction risk test" (Robinson, 1999) was utilized to capture the most widely accepted conceptualization of workaholism: compulsive dependence to work and perfectionist tendencies. The measure consists of 25 self-report items where participants are asked to respond using a 4-point summated rating scale with anchors of "not at all true" and "completely true." Internal consistency (Cronbach's coefficient alpha) was estimated at 0.88, and test-retest reliability was measured at 0.83 (Robinson, Post, & Khakee, 1992). Further, split-half reliability using three data sets resulted in a Spearman-Brown split-half coefficient of 0.85 (Robinson & Post, 1995). Overall, assessments of face, criterion-related, content, and concurrent validity have indicated that the scale consistently embodies the intended construct of workaholism (Robinson, 1999; Robinson & Post, 1995).

Included in the hypothesized model are the key role stressors of role overload, role conflict, and role ambiguity. Role overload was measured using a scale from Beehr et al. (1976). Role conflict and role ambiguity were assessed using scales based on the measures developed by Rizzo, House, and Lirtzman (1970). All three measures utilized a 7-point Likert scale (anchored by "strongly agree" and "strongly disagree"), requesting participants to respond to statements relative to their work environment. All three of these measures have undergone psychometric analyses and are widely used and accepted (Fogarty et al., 2000; Jackson & Schuler, 1985). All three role

measures have been utilized in recent exhaustion studies involving public accountants (Almer & Kaplan, 2002; Fogarty et al., 2000; Sweeney & Summers, 2002).

For the workload variable, subjects were asked to indicate how many hours they had worked per week during the prior month. This question is consistent with that utilized by Sweeney and Summers (2002) in their study examining the effects of busy-season workload on public accountants' burnout.

ANALYSES AND RESULTS

Descriptive Statistics

Table 2 presents means, standard deviations, and scale reliabilities (Cronbach's alpha) for the model variables. The mean score for exhaustion (2.60) represents a relatively high level of exhaustion when compared with samples from other high-stress professions such as law enforcement (Gaines & Jermier, 1983), nursing (Maslach & Jackson, 1986), and human services (Lee & Ashforth, 1993), but is lower than both the pre-busy-season (2.97) and busy-season (3.46) levels of exhaustion measured by Sweeney and Summers (2002) in their study of public accountants.

Although the data for the current study were collected during May and November, respectively, for the regional and national firms, average hours

Table 2. Descriptive Statistics for Model Variables and Estimates of Internal Consistency for Scale Items.

Variable	Overall		Firm A		Firm B		Alpha
	Mean	SD	Mean	SD	Mean	SD	
Exhaustion	2.60	0.97	2.76	1.04	2.29	0.77	0.92
Neuroticism	2.93	0.99	3.01	1.06	2.79	0.83	0.85
Hardiness	3.01	0.23	2.98	0.23	3.06	0.21	0.75
Type-A	5.96	1.10	5.95	1.16	5.98	1.01	0.61
Workload	2.06	0.36	2.05	0.37	2.08	0.34	0.82
Role overload	3.78	1.07	3.88	1.02	3.62	1.16	0.63
Role conflict	3.95	1.18	4.18	1.07	3.54	1.27	0.86
Role ambiguity	3.46	1.14	3.67	1.16	3.08	1.00	0.89
Hours	41.60	8.34	41.28	7.29	42.18	10.04	N/A

Note: *n* = 112.

Table 3. Correlations among Model Variables.

	EXH	NEU	HAR	TPA	WOR	RLO	RLC	RLA	HRS
EXH	1.00								
NEU	0.28**	1.00							
HAR	-0.42**	-0.37**	1.00						
TPA	0.08	0.40**	0.04	1.00					
WOR	0.21*	0.37**	-0.23*	0.38**	1.00				
RLO	0.48**	0.11	0.02	0.13	0.10	1.00			
RLC	0.39**	0.22*	-0.17	0.17	0.33**	0.41**	1.00		
RLA	0.42**	0.15	-0.34**	-0.14	0.13	0.24*	0.41**	1.00	
HRS	-0.22*	-0.02	0.04	0.17	0.12	-0.06	0.06	-0.14	1.00

Note: $n = 112$. EXH = exhaustion (Maslach & Jackson, 1986); NEU = neuroticism (Saucier, 1994); HAR = hardiness (Bartone et al., 1989); TPA = Type-A behavior (Bortner, 1969); WOR = Workaholism (Robinson, 1999); RLO = role overload (Beehr, Welsh, & Taber, 1976); RLC = role conflict (Rizzo et al., 1970); RLA = role ambiguity (Rizzo et al., 1970); and HRS = workload in average weekly hours.

* $p < .05$, ** $p < .01$.

worked per week were fairly consistent between firms (mean = 42.18). The mean weekly workload was lower than that reported by Sweeney and Summers (2002) for public accountants during the busy season (mean = 62.7) and just prior to the busy season (48.9).

Correlations for all model variables are found in Table 3. As expected, exhaustion was positively correlated with neuroticism, workaholism, and the role variables, and negatively correlated with hardiness. Exhaustion was not significantly correlated with Type-A behavior. Contrary to expectations, exhaustion was negatively correlated ($p < .05$) with workload outside the busy season.

Tests of Hypotheses

The hypotheses were tested using path analysis following the theoretical model presented in Fig. 1.⁵ The analyses that follow were conducted in two ways. First, the analyses were performed using a simple average of the observed variables to create the measured constructs. For example, the eight items representing role conflict are averaged into a single variable (after adjusting reverse scored items). Two, the analyses were performed using factors (measured constructs) created through factor analysis. For example, the eight items representing role conflict were factor analyzed to determine

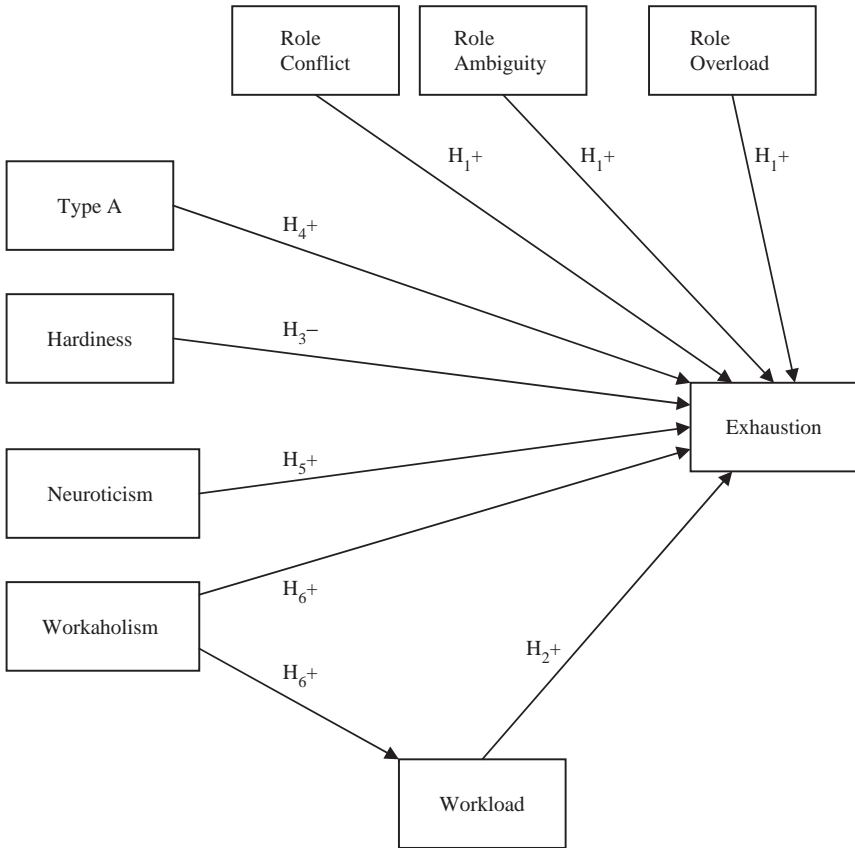


Fig. 1. Theoretical Model.

the shared variance of the resulting primary factor. During this analysis a varimax rotation was performed and the output variable was created using a Bartlett Score estimation. This procedure allows for the maximum shared variance to be captured in the measured construct inserted into the path analysis. While both methods were used to create the path analysis, only the simple average method is reported as the results were statistically equivalent. All significant relationships reported were also found in the factored analysis and none of the insignificant relations reported here were significant with the factored analysis.

The initial model produced via path analysis followed the theoretical model in Fig. 1 but allowed for correlation among the four personality

variables; Type-A, neuroticism, workaholism, and hardiness. The model also allowed for correlation among the three role variables. Overall, the model is fit reasonably well with a CFI of .88 and a RMSEA of .10. This model is shown in Fig. 2.

The first two hypotheses examined relationships between contextual variables and accountants' exhaustion. H₁ predicted a positive relationship among role stressors and exhaustion, and is partially supported. Role

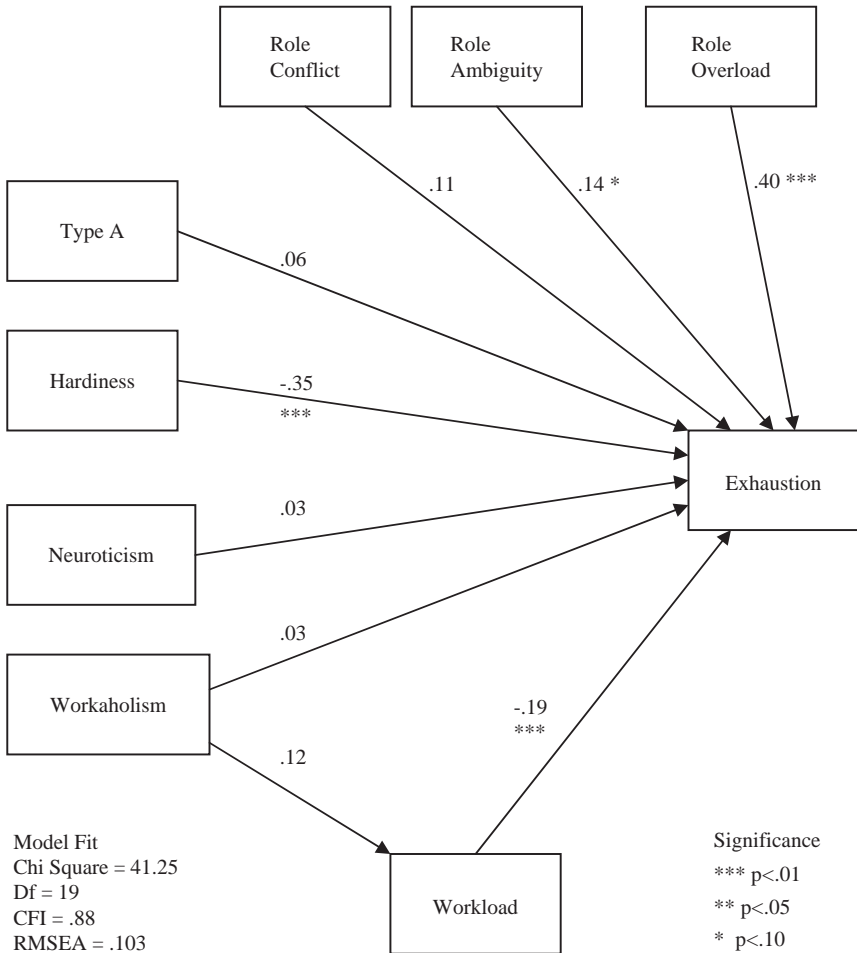


Fig. 2. Tests of Hypotheses Full Model.

overload was positively related to exhaustion and significant ($p < .0001$), while role ambiguity was marginally significant ($p = .063$) and role conflict was insignificant.

The predicted relationship in H_2 between workload and exhaustion was significant but the sign of the relationship was unexpectedly negative. A potential explanation for the negative relationship is that for the majority of subjects, data collection occurred shortly after busy season. During the immediate post-busy season, those public accountants with the highest busy-season workload may have experienced the greatest decreases in workload, but their exhaustion may have been remained relatively high. The remaining hypotheses were concerned with the influence of individual personality variables on exhaustion. H_3 predicts that accountants high in hardiness will experience less exhaustion. This hypothesis is supported by the path analysis, as hardiness is highly significant ($p < .0001$) and in the predicted direction.

H_4 was not supported. Public accountants who exhibited greater Type-A tendencies did not experience greater exhaustion. This result is in contrast to some prior research that found a relationship between Type-A and stress with public accounting subjects (Choo, 1986; Haskins et al., 1990).⁶ A potential explanation for the lack of significance is that personality is multifaceted, and some traits intercorrelate so highly that they may be considered measures of an equivalent construct (Judge et al., 2002; Watson & Clark, 1984). In contrast to prior accounting stress research that utilized only a single personality trait, this study examined multiple personality traits concurrently, and the Type-A trait may have been redundant or overlapping with other traits included in the model.

The personality trait of neuroticism, although correlated with exhaustion, was insignificant in the path analysis, leading to a rejection of H_5 . H_6 was also not supported, as the influence of workaholism on workload (hours) was insignificant. The combined direct and indirect effects of workaholism on exhaustion are also insignificant.

Additional Analyses

Given the relatively small sample used to test the theoretical model and the number of hypothesis that were unsupported by the data, some exploratory analysis was performed to examine the strength of the model in the absence of the insignificant hypothesized relationships. As this exploratory analysis moves only toward a reduction of the theoretical model and does not

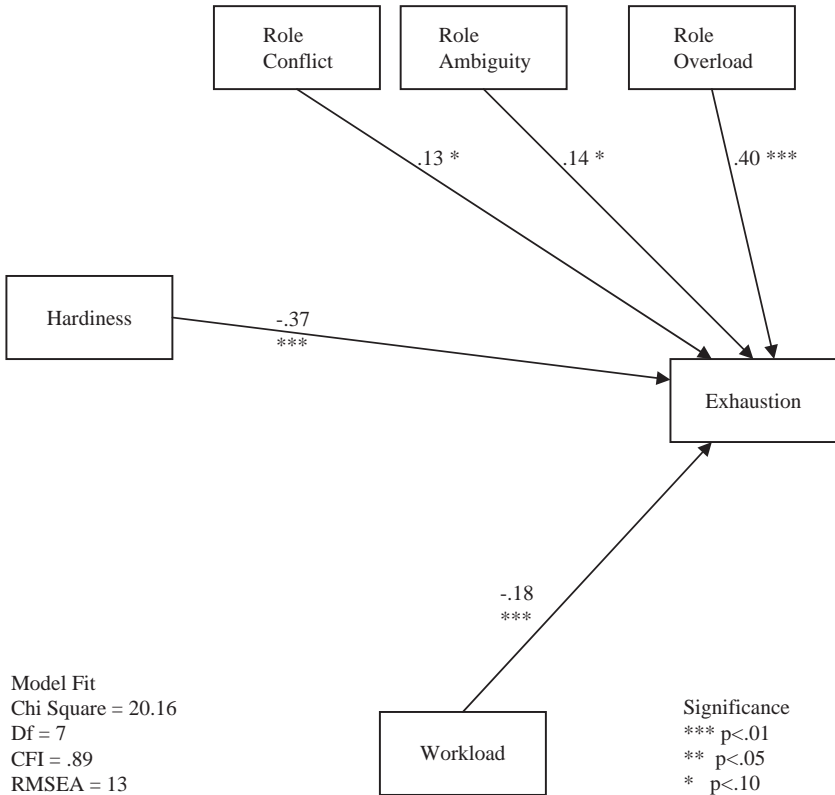


Fig. 3. Test of Hypotheses Reduced Model.

propose any additional relationships, it should not be construed as a “fishing expedition.” The reduced model in Fig. 3 shows the influence of the role variables, hardiness, and workload on exhaustion.

As indicated in Fig. 3, the model fit is marginal with a CFI of .89 and a RMSEA of .13. The influence of role overload, role conflict, hardiness, and workload are largely unchanged. The influence of role ambiguity is somewhat stronger at .13 and becomes marginally significant at $p = .10$. The number of estimated parameters in the reduced model is 21, which is within an acceptable range for path analysis on a sample of this size.

As the introduction of hardiness’s influence on exhaustion is the most salient addition to the literature from this research, replacing the overall

measure of hardiness in the path model with its three components (commitment, challenge, and control) may shed light on the dimensions of hardiness that are most salient to public accountants' exhaustion. As shown in Fig. 4, the results indicate that commitment and challenge are significantly and inversely related to exhaustion in public accountants. The control component was not significant. This indicates that an individual with relatively high levels of challenge and commitment may be more able to withstand the stressful demands of the public accounting environment.

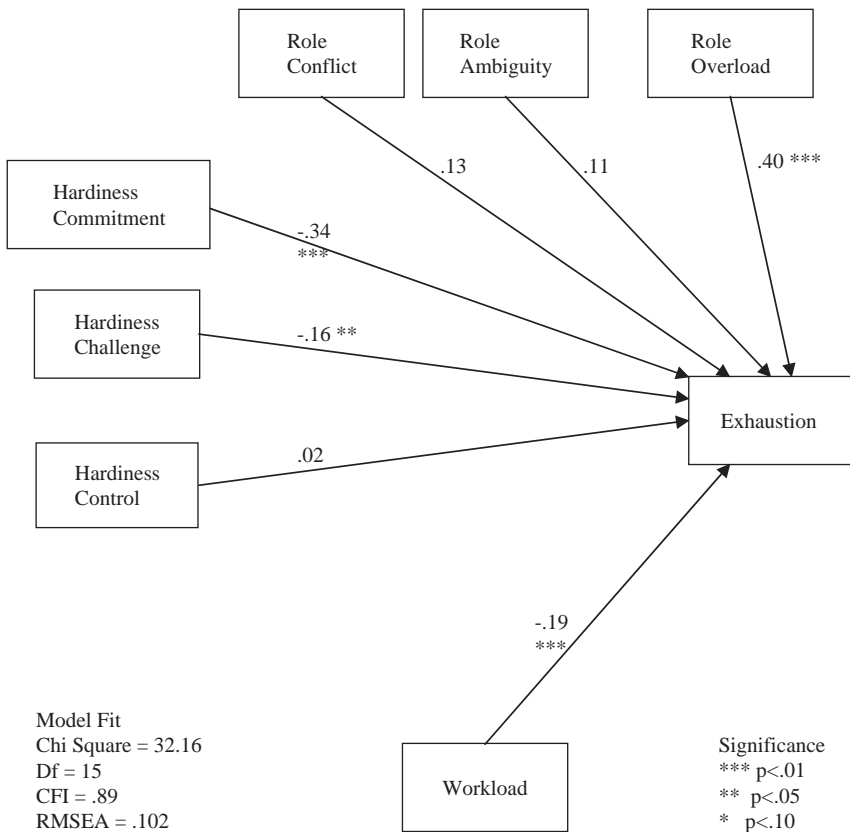


Fig. 4. Test of Hypotheses Expanded Hardiness Model.

DISCUSSION AND CONCLUSION

The stress variable of exhaustion has significant costs in terms of employees' physical and psychological well being, as well as negative organizational consequences (Halbesleben & Buckley, 2004). Prior accounting research has focused on the relationship between contextual factors and public accountants' exhaustion. This study examines both the effect of contextual variables, such as workload and role stressors, and individual personality differences on public accountants' exhaustion. The advantage of examining multiple personality traits is that overlapping constructs can be identified, as well as variables that contribute uniquely to public accountants' exhaustion.

An important contribution of this research is the finding that public accountants low in hardiness may be less resistant to exhaustion. A supplemental analysis indicated that the hardiness dimensions of commitment and challenge were significantly and inversely related to exhaustion. By revealing the relationship between hardiness and exhaustion, management may be able to implement appropriate treatment strategies, such as offering employees the opportunity to work on a flextime arrangement (Almer & Kaplan, 2002).

There are several limitations of the current study. First, as with all studies utilizing cross-sectional survey data, the potential for common-method bias is present. Second, the study relied upon responses from professionals in two firms who were not randomly selected, which may impair generalizability. Furthermore, the firms in the study may not be representative of the population of public accounting firms, and caution should be exercised in extrapolating the results of this study to accountants in large, international accounting firms. Finally, the cross-sectional, survey methodology utilized in the study does not rule out the possibility that other, unmeasured personality or contextual variables may impact exhaustion.

This study is the first to examine public accountants' hardiness in the context of exhaustion, and the results indicate that this personality trait may have an effect on the ability of public accountants to be successful in coping with high-stress environments. Researchers examining stress in the domain of public accounting should consider including the hardiness construct in their models, as well as investigating its impact on other important variables, such as job performance. Based on the results of this study and those of Sweeney and Summers (2002), we suggest that future research differentiate between the measurement of public accountants' exhaustion during the busy season versus outside of the busy season, as these time periods may produce different relationships between exhaustion and workload.

NOTES

1. Throughout the chapter's narrative, the term "emotional exhaustion" has been shortened to "exhaustion" to improve readability. This is consistent with the current literature (Maslach et al., 2001).

2. Moore (2000a) argues that burnout research would be best served by isolating the critical emotional exhaustion component from depersonalization and reduced personal accomplishment.

3. The "Big Five" personality dimensions include extraversion, agreeableness, conscientiousness, neuroticism, and intellect or openness (Costa & McCrae, 1985; Mount & Barrick, 1995).

4. Sweeney and Summers (2002) collected longitudinal data just prior to the start of busy season (November–December) and at the end of busy season (March).

5. Due to the limited sample size, we utilized path analysis in lieu of structural equation modeling. A limitation of path analysis is its inability to use maximum likelihood optimization techniques to extract the shared variance of the observed variables representing latent construct, while at the same time estimating the relationships among latent variables. The use of a direct effects model may not adequately express the possible relationships in the data, but the limited sample size limits exploration.

6. Kelley and Margheim (1990) hypothesized a moderating effect of Type-A behavior on the effects of time budget pressures in auditors. The authors ultimately found no such relationship.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the helpful comments of Bernie Wong-On-Wing, Bill Stammerjohan, and other participants in the accounting research workshop at Washington State University.

REFERENCES

- Allen, J., & Mellor, D. (2002). Work context, personal control, and burnout amongst nurses. *Western Journal of Nursing Research, 24*(8), 905–917.
- Almer, E. D., & Kaplan, S. E. (2002). The effects of flexible work arrangements on stressors, burnout, and behavioral outcomes in public accounting. *Behavioral Research in Accounting, 14*, 1–34.
- Babakus, E., Cravens, D. W., Johnston, M., & Moncrief, W. C. (1999). The role of emotional exhaustion in sales force attitude and behavior relationships. *Journal of the Academy of Marketing Science, 27*(1), 58–70.
- Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth.
- Bakker, A. B., Van Der Zee, K. I., Lewig, K. A., & Dollard, M. F. (2006). The relationship between the Big Five personality factors and burnout: A study among volunteer counselors. *The Journal of Social Psychology, 146*(1), 31–50.

- Bartone, P. T., Ursano, R. J., Wright, K. M., & Ingraham, L. H. (1989). The impact of a military air disaster on the health of assistance workers. *The Journal of Nervous and Mental Disease*, 177(6), 317–328.
- Beehr, T. A., Welsh, J., & Taber, T. (1976). Relationship of stress to individually and organizationally valued states: Higher order needs as moderator. *Journal of Applied Psychology*, 61(1), 41–47.
- Bortner, R. W. (1969). A short rating scale as a potential measure of pattern a behavior. *Journal of Chronic Diseases*, 12, 87–91.
- Cano-García, F. J., Padilla-Munoz, E. M., & Carrasco-Ortiz, M. A. (2005). Personality and contextual variables in teacher burnout. *Personality and Individual Differences*, 38, 929–940.
- Choo, F. (1986). Job stress, job performance, and auditor personality characteristics. *Auditing: A Journal of Practice and Theory* (Spring), 17–34.
- Collins, K. M., & Killough, L. N. (1992). An empirical examination of stress in public accounting. *Accounting, Organizations and Society*, 17, 535–547.
- Constantini, A., Solano, L., Di Napoli, R., & Bosco, A. (1997). Relationship between and risk of burnout in a sample of 92 nurses working in oncology and AIDS wards. *Psychotherapy and Psychosomatics*, 66(2), 77–82.
- Cordes, C. L., & Dougherty, T. W. (1993). A review and integration of research on job burnout. *Academy of Management Review*, 18, 621–656.
- Cordes, C. L., Dougherty, T. W., & Blum, M. (1997). Patterns of burnout among managers and professionals: A comparison of models. *Journal of Organizational Behaviour*, 18, 685–701.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual*. Psychological Assessment Resources.
- Demetriou, A., Kyriakides, L., & Avraamidou, C. (2003). The missing link in the relations between intelligence and personality. *Journal of Research in Personality*, 37, 547–581.
- de Vries, J., & van Heck, G. L. (2000). Persoonlijkheid en emotionele uitputting. Een overzicht van de literatuur. *Gedrag & Gezondheid: Tijdschrift voor Psychologie & Gezondheid*, 28(3), 90–105.
- Eysenck, H. J., & Eysenck, S. G. B. (1963). *Eysenck Personality Inventory, Form A*. San Diego: Educational and Industrial Testing Service.
- Fassel, D. (1990). *Working ourselves to death: The high cost of workaholism, the rewards of recovery*. San Francisco, CA: Harper Collins Publishers.
- Fisher, R. T. (2001). Role stress, the Type A behavior pattern, and external auditor job satisfaction and performance. *Behavioral Research in Accounting*, 13, 143–170.
- Fogarty, T. J., Singh, J., Rhoads, G. K., & Moore, R. K. (2000). Antecedents and consequences of burnout in accounting: Beyond the role stress model. *Behavioral Research in Accounting*, 12, 31–67.
- Funk, S. C. (1992). Hardiness: A review of theory and research. *Health Psychology*, 11(5), 335–345.
- Gaines, J., & Jermier, J. M. (1983). Emotional exhaustion in a high-stress organization. *Academy of Management Journal*, 26, 567–586.
- Golembiewski, R. T., Boudreau, R. A., Munzenrider, R. F., & Luo, H. (1996). *Global burnout: A worldwide pandemic explored by the phase model*. Greenwich, CT: JAI Press.
- Goodard, R., Patton, W., & Creed, P. (2004). The importance and place of neuroticism in predicting burnout in employment service case managers. *Journal of Applied Social Psychology*, 34, 282–296.

- Halbesleben, J. R. B., & Buckley, M. R. (2004). Burnout in organizational life. *Journal of Management*, 30(6), 859–879.
- Haskins, M., Baglioni, A., & Cooper, C. (1990). An investigation of the sources, moderators, and psychological symptoms of stress among audit seniors. *Contemporary Accounting Research*, 6, 361–385.
- Houkes, I., Janssen, P. P. M., de Jonge, J., & Bakker, A. B. (2003). Specific determinants of intrinsic work motivation, emotional exhaustion and turnover intention: A multisample longitudinal study. *Journal of Occupational and Organizational Psychology*, 76, 427–450.
- Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Decision Processes*, 36, 16–78.
- Jamal, M. (1990). Relationship of job stress and Type-A behavior to employee's job satisfaction, organizational commitment, psychosomatic health problems, and turnover motivation. *Human Relations*, 43, 727–738.
- Judge, T. A., Erez, A., Thoresen, C. J., & Bono, J. E. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?. *Journal of Personality and Social Psychology*, 83(3), 693–710.
- Kelley, T., & Margheim, L. (1990). The impact of time budget pressure, personality, and leadership variables on dysfunctional auditor behavior. *Auditing: A Journal of Practice and Theory*, 9(2), 21–42.
- Kim, J., Shin, K., & Umbreit, T. (2007). Hotel job burnout: The role of personality characteristics. *International Journal of Hospitality Management*, 26(2), 421–434.
- Kobasa, S. C. (1979). Stressful life events, personality, and health: An inquiry into hardiness. *Journal of Personality and Social Psychology*, 37, 1–11.
- Kobasa, S. C., Maddi, S. R., & Courington, S. (1981). Personality and constitution as mediators in the stress-illness relationship. *Journal of Health and Social Behavior*, 22, 368–378.
- Koeske, G. F., & Koeske, R. D. (1989). Construct validity of the MBI: A critical review and reconceptualization. *Journal of Applied Behavioral Science*, 25, 131–144.
- Lampe, J. C. (2004). Alternative personality measurements: Commentary on accounting information systems research using opportunities using personality type theory and the Myers-Brigg Type indicator. *Journal of Information Systems*, 18(Spring), 21–34.
- Law, D. W. (2005). Interactive organizational commitment and hardiness in public accountants' turnover. *Managerial Auditing Journal*, 20(4), 383–393.
- Lee, R. T., & Ashforth, B. E. (1993). A further examination of managerial burnout: Toward an integrated model. *Journal of Organizational Behavior*, 14, 3–20.
- Lee, R. T., & Ashforth, B. E. (1996). A meta-analytic examination of the correlates of the three dimensions of job burnout. *Journal of Applied Psychology*, 2, 123–133.
- Leiter, M. P., & Durup, J. M. (1996). Work, home, and in between: A longitudinal study of spillover. *Journal of Applied Behavioral Science*, 32, 29–47.
- Maddi, S. R., & Kobasa, S. C. (1984). *The hardy executive: Health under stress*. Homewood, IL: Dow Jones-Irwin.
- Martin, J. J., Kelley, B., & Eklund, R. C. (1999). A model of stress and burnout in male high school athletic directors. *Journal of Sport and Exercise Psychology*, 21(3), 280–294.
- Maslach, C. (1982). *Burnout: The cost of caring*. Englewood Cliffs, NJ: Prentice-Hall.
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior*, 2, 99–113.

- Maslach, C., & Jackson, S. E. (1986). *Maslach burnout inventory manual*. Palo Alto, CA: Consulting Psychologists Press.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach Burnout Inventory Manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Maslach, C., & Leiter, M. (1997). *The truth about burnout*. San Francisco, CA: Jossey-Bass Publishers.
- Maslach, C., Schaufeli, W., & Leiter, M. (2001). Job burnout. *Annual Review of Psychology*, 52, 397–422.
- Matthews, K. A. (1982). Psychological perspectives on the type A behavior pattern. *Psychological Bulletin*, 91, 293–323.
- Mayes, B. T., Johnson, T. W., & Sadri, G. (2000). Personality, job levels, job stressors, and their interaction as predictors of coping behavior. *Psychological Reports*, 87, 61–81.
- McCranie, E. W., Lambert, V. A., & Lambert, C. E. (1987). Work stress, hardiness, and burnout among hospital staff nurses. *Nursing Research*, 36(6), 374–378.
- Moore, J. (2000a). Why is this happening? A causal attribution approach to work exhaustion consequences. *The Academy of Management Review* (April), 1–16.
- Moore, J. (2000b). One road to turnover: An examination of work exhaustion in technology professionals. *MIS Quarterly*, 24, 141–168.
- Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resource management. In: G. R. Ferris & Rollins (Eds), *Research in personnel and human resources management* (Vol. 13, pp. 153–200). Greenwich, CT: JAI Press.
- Nagy, S., & Davis, L. G. (1985). Burnout: A comparative analysis of personality and environmental variables. *Psychological Reports*, 57, 1319–1326.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Pierce, C. M., & Malloy, G. N. (1990). Psychological and biographical differences between secondary school teachers experiencing high and low levels of burnout. *British Journal of Educational Psychology*, 60, 37–51.
- Price, V. A. (1982). *Type A behavior: A model for research and practice*. New York, NY: Academic Press.
- Rich, V., & Rich, A. (1987). Personality hardiness and burnout in female staff nurses. *Image*, 19, 63–66.
- Rizzo, J., House, J., & Lirtzman, S. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly*, 15, 150–163.
- Robinson, B. E. (1999). The work addiction risk test: Development of a tentative measure of workaholism. *Perceptual and Motor Skills*, 88, 199–210.
- Robinson, B. E., & Post, P. (1995). Split-half reliability of the work addiction risk test: Development of a measure of workaholism. *Psychological Reports*, 76, 1226.
- Robinson, B. E., Post, P., & Khakee, J. F. (1992). Test-retest reliability of the Work Addiction Risk Test. *Perceptual and Motor Skills*, 74, 926.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment*, 63(3), 506–516.
- Senetra, P. (1980). Role conflict, role ambiguity, and organizational climate in a public accounting firm. *The Accounting Review*, 55(October), 594–603.
- Sheard, M., & Golby, J. (2007). Hardiness and undergraduate academic study: The moderating role of commitment. *Personality and Individual Differences*, 43(3), 579–588.

- Shirom, A. (1989). Burnout in work organizations. In: C. L. Cooper & I. Robertson (Eds), *International review of industrial and organizational psychology* (pp. 25–48). New York, NY: Wiley.
- Smith, K. J., Everly, G. S., & Johns, T. R. (1993). Stress arousal in the stressor-to-illness process. *Contemporary Accounting Research*, 9(Spring), 432–449.
- Snead, K., & Harrell, A. (1991). The impact of psychological factors on the job satisfaction of senior auditors. *Behavioral Research in Accounting*, 3, 85–96.
- Spector, P. E., & O'Connell, B. J. (1994). The contribution of personality traits, negative affectivity, locus of control and Type A to the subsequent reports of job stressors and job strains. *Journal of Occupational and Organizational Psychology*, 67, 1–11.
- Spence, J. T., & Robbins, A. S. (1992). Workaholism: Definition, measurement, and preliminary results. *Journal of Personality Assessment*, 58(1), 160–178.
- Sweeney, J. T., & Summers, S. L. (2002). The effect of the busy season workload on public accountants' job burnout. *Behavioral Research in Accounting*, 14, 223–245.
- Tokar, D. M., Fischer, A. R., & Subich, L. M. (1998). Personality and vocational behavior: A selective review of the literature, 1993–1997. *Journal of Vocational Behavior*, 53, 115–153.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465–490.
- Weick, K. E. (1983). Stress in accounting systems. *The Accounting Review* (April), 350–369.
- Witt, L. A., Andrews, M. C., & Carlson, D. S. (2004). When conscientiousness isn't enough: Emotional exhaustion and performance among call center customer service representatives. *Journal of Management*, 30(1), 149–160.
- Wright, T. A., & Cropanzano, R. (1998). Emotional exhaustion as a predictor of job performance and voluntary turnover. *Journal of Applied Psychology*, 30, 1570–1598.
- Zellars, K., Perrewe, P., & Hochwarter, A. R. (2000). Burnout in health care: The role of the five factors of personality. *Journal of Applied Social Psychology*, 30, 1570–1598.

A RESEARCH NOTE ON THE EFFECTS OF FINANCIAL AND NONFINANCIAL MEASURES IN BALANCED SCORECARD EVALUATIONS

Kip R. Krumwiede, Tim V. Eaton, Monte R. Swain
and Dennis Eggett

ABSTRACT

Schiff and Hoffman (1996) found evidence that nonfinancial measures explain more of the variance in evaluations that focus on individual retail department managers while financial measures explain more variance in evaluations of the overall department. These findings are consistent with Attribution Theory, which holds that evaluators of performance ascribe cause to individual or environmental factors as they make judgments. This study expands this research by being the first to examine whether financial and nonfinancial measures affect multidivisional balanced scorecard performance evaluations differently when the focus of the evaluation is on the individual division president versus when the focus is on the overall division. The results of this study suggest that when evaluating individual performance, nonfinancial measures clearly affect the performance evaluations more than financial measures. When the

Advances in Accounting Behavioral Research, Volume 11, 155–177

Copyright © 2008 by Emerald Group Publishing Limited

All rights of reproduction in any form reserved

ISSN: 1475-1488/doi:10.1016/S1475-1488(08)11007-9

focus is on the division, the influence of nonfinancial and financial measures is not differentiated. Additionally, the results suggest that the participants perceived nonfinancial measures to be more controllable than financial measures.

INTRODUCTION

This study examines how financial and nonfinancial measures affect balanced scorecard (BSC) performance evaluations. This question is important because prior research suggests that evaluations at all organizational levels increasingly include a significant number of nonfinancial measures (Ittner, Larcker, & Rajan, 1997; Said, HassabElnaby, & Wier, 2003; Chow & Van Der Stede, 2006). Additionally, recent reports suggest that approximately 50 percent of all Fortune 1,000 companies and 40 percent of European companies now use some version of the BSC (Silk, 1998; Gumbus & Lyons, 2002; Brewer, 2002; Salterio & Webb, 2003). However, while the BSC has garnered enthusiastic support in some companies, it can also be costly to develop and can lead to undesirable performance evaluations and unsuccessful implementations (Lipe & Salterio, 2000; Ittner, Larcker, & Meyer, 2003). Therefore, more needs to be understood about how BSC evaluation decisions are made (Ittner & Larcker, 1998; Ittner et al., 2003). Specifically, because managers will focus their efforts on processes for which they are evaluated, understanding how evaluators use financial and nonfinancial information in their evaluations is important to both the design and the implementation of effective scorecard systems (Hopwood, 1972; Kaplan & Norton, 1992; Banker, Chang, & Pizzini, 2004).

Schiff and Hoffman (1996) found that variance in departmental judgments was explained more by financial (FIN) measures than by nonfinancial measures while nonfinancial measures explained more of the variance in manager evaluations. They suggest that these findings are consistent with Attribution Theory (AT), which holds that evaluators ascribe cause to individual or environmental factors as they make judgments. However, prior BSC studies have focused on individual manager evaluations and the common measures bias (Lipe & Salterio, 2000; Libby, Salterio, & Webb, 2004; Roberts, Albright, & Hibbets, 2004; Banker et al., 2004; Dilla & Steinbart, 2005). This study extends the results of Schiff and Hoffman to a BSC context. Specifically, it addresses the question of whether evaluators using BSC information will use financial and nonfinancial measures differently depending on the focus of the evaluation.

The contribution of the present study is that it is the first to provide evidence that financial and nonfinancial measures affect BSC performance evaluations differently when the focus of the evaluation is on the individual division president versus the overall division. The results of this study suggest that in a multidivisional BSC context, nonfinancial measures clearly affect evaluations more than FIN measures when the focus is on individual performance. When the focus is on the division, the influence of nonfinancial and FIN measures is not differentiated. Additionally, the results suggest that the participants perceived nonfinancial measures to be more controllable than FIN measures. One of the causal dimensions of AT is controllability, which may help explain why the participants were influenced so heavily by the nonfinancial measures.

The following sections develop the research hypotheses through an examination of the relevant literature, describe the experimental research method, report the results, and discuss the study's contributions and limitations.

RESEARCH HYPOTHESES

Nonfinancial Performance Measures

A fundamental component of the BSC approach is the inclusion of financial and nonfinancial performance measures closely linked to the competitive strategies of the organization (Kaplan & Norton, 1992, 1993, 1996, 2001). However, the use of nonfinancial performance measures to establish and manage strategic objectives should be considered carefully when these measures are used to evaluate performance. Ittner and Larcker (1998, p. 228) underscored this point with their statement that an "important question is whether the same measures or scorecard used to develop strategic priorities and monitor strategic actions should be used to evaluate managerial performance." However, recent experimental research has focused on the "common measures bias" first reported in Lipe and Salterio's (2000) seminal study, which found that performance evaluations of managers are affected more by common measures than by unique measures.

Lipe and Salterio (2000) noted that measures that are common across business units often tend to be the financial indicators that lag behind performance, while the unique measures are often nonfinancial measures that lead (i.e., precede) financial results. Subsequent BSC experimental studies examined factors that mitigate the common measures bias, including

process accountability and third-party assurance reports (Libby et al., 2004), disaggregating information (Roberts et al., 2004), articulation of the linkage between performance measures and business strategy (Banker et al., 2004), and task experience (Dilla & Steinbart, 2005). However, these studies do not directly address the impact of FIN measures versus nonfinancial measures nor the issue of target focus (i.e., all were evaluations of the individual managers). Ghosh (2005) found evidence that nonfinancial measures are perceived to be more controllable than FIN measures and, therefore, have more impact on the “outcome effect” in performance evaluation decisions. Chow and Van Der Stede (2006) surveyed 128 manufacturing firms in the United States and Europe and found that nonfinancial measures are seen by manufacturing managers as providing the greatest encouragement for risk taking and innovation but they also found that evaluations of manufacturing performance emphasize FIN measures over nonfinancial measures (49 versus 30 percent, respectively). However, this study appears to focus on evaluations of organizational performance rather than on the performance of individual managers.

The context of the present study is a BSC system with a set of financial and nonfinancial measures that are common across two divisions. This design is necessary to directly observe how financial and nonfinancial measures affect performance evaluation decisions of two different targets, the division and the division manager.

Target Focus of Evaluation

The present study examines whether financial and nonfinancial measures affect the BSC evaluation scores differently depending on the target focus. The results of Schiff and Hoffman (1996) suggest that evaluators of organizations rely on financial and nonfinancial measures differently than evaluators of the managers of those organizations. They found that evaluations of departmental performance are associated more with causes *external* to the individual manager (e.g., competition, economic environment, or luck) and that evaluators weight *financial* measures more heavily than nonfinancial factors. In contrast, evaluations of departmental managers were found to be associated more with causes *internal* to the individual manager (e.g., intellect, effort, or skill) and that evaluators weight *nonfinancial* measures more heavily. In discussing the application of the results of Schiff and Hoffman’s study to a BSC context, Ittner and Larcker (1998, p. 229) state that a “potential area of research is examining the

relative value of different types of scorecards and performance measures for different purposes.”

Schiff and Hoffman link their paper to AT. According to AT, individuals making judgments attempt to establish causal explanations for events in arriving at these judgments (Kaplan & Reckers, 1985; Martinko, 1995). AT proposes that evaluators of performance may ascribe these causal explanations to the individual being evaluated or to environmental (situational) factors (Martinko, 1995). Ascribing cause to the individual is defined as an internal attribution, in contrast to ascribing cause to the situation (or environment), which is defined as an external attribution (Kaplan & Reckers, 1985, 1991; Martinko, 1995). AT has been applied in different accounting contexts, including accounting for tax evasion (Kaplan, Reckers, & Roark, 1988), auditor performance evaluation (Kaplan & Reckers, 1985, 1991), and auditor planning judgments (Todd, Houston, & Peters, 2001). Kaplan and Reckers (1985) identified an AT phenomenon for performance evaluation within an audit environment. They reported that evaluators demonstrate a greater tendency to take action against the individual auditor when underlying events are seen as having higher internal attribution.

Very limited research has examined AT in management performance evaluation. Schiff and Hoffman (1996) posited that if AT is reflected in how performance measures are used, it may then be inferred that evaluators will associate some measures with causes that are internal to the individual, while other measures will be judged as more or less beyond the individual's control. Their findings support this proposition and emphasize controllability, one of the causal dimensions of AT (Weiner, 1985; Kent & Martinko, 1995).

The controllability principle as it applies to evaluation of managers holds that managers should be held responsible (evaluated) only for that which they control (Antle & Demski, 1988, pp. 700–701; Zimmerman, 1995, p. 170; Ghosh, 2005). Ghosh (2005) finds that nonfinancial measures tend to have higher perceived controllability than FIN measures because the latter are too aggregated and may not provide good information about individual performance. Nonfinancial measures are regarded as better signals of actions that managers can take to improve overall performance. FIN measures are generally regarded as lagging measures of performance, while nonfinancial measures are often leading measures of financial performance (Kaplan & Norton, 1996; Banker et al., 2004; Bryant, Jones, & Widener, 2004). Ghosh also found that the outcome effect, which occurs when outcome knowledge influences the evaluator's assessment of a manager, increases as the perceived controllability of the measures increases. These results are consistent with prior research (DeNisi, 1996; Tan & Lipe, 1997).

AT suggests that evaluators will attribute the cause of good or poor performance to either internal or external factors. The results of Schiff and Hoffman (1996) (hereafter, S&H) suggest that evaluators of division presidents will ascribe a stronger link between the individual's performance and internal factors, measured best by nonfinancial measures that are more closely attributed to individual manager actions. The results of Ghosh (2005) also support this prediction since nonfinancial measures are also expected to be perceived as more controllable than FIN measures because FIN measures may not provide information on the individual manager's performance and are very aggregated (Ittner & Larcker, 1998). Because the outcome effect increases as the perceived controllability of the measures increases, the more controllable nonfinancial measures, rather than the less controllable FIN measures, should have more effect on the judgments of those who evaluate the division presidents. Thus, in a BSC context, evaluators of division presidents are expected to be affected more by nonfinancial measures in their evaluations than by FIN measures.

The effect of nonfinancial versus FIN measures on evaluators of division performance is harder to differentiate. S&H found that because of higher external attribution, evaluators of the department performance weight FIN measures more highly than evaluators of the department manager. This suggests that in a BSC context, division evaluators will be influenced more by FIN measures. On the other hand, Ghosh's (2005) study suggests that nonfinancial measures will be perceived to have higher controllability and be more indicative of future division performance than will FIN measures. However, Ghosh only considered a context in which individual store managers were evaluated rather than the entire store. Following S&H, when evaluators focus their evaluation on division performance, they are expected to be influenced more by FIN measures than by nonfinancial measures.

Based on the preceding discussion, the following hypotheses will be tested (alternative form):

H₁. Evaluators who focus specifically on the division *president* rather than on the division operating unit will be affected more heavily by *nonfinancial* measures than by financial measures in their decisions.

H₂. Evaluators who focus specifically on the *division* operating unit rather than on the division president will be affected more heavily by *financial* measures than by nonfinancial measures in their decisions.

METHODOLOGY

Experimental Case

This study uses an experimental case approach based on a fictitious company entitled Work-Right with two divisions, Excavation Division and Commercial Division.¹ To enhance external validity, the Work-Right case was adapted from a large manufacturing company's actual scorecard.² The participants were told to assume the role of the CEO of Work-Right. Participants were split into two groups, either to evaluate the performance of the two division presidents or to evaluate the performance of the two divisions. Those who evaluated the presidents of the two divisions were told that their decisions would be used for bonus and promotion considerations. Evaluators of the two divisions were told that their assessments would be used for resource allocation decisions. All participants were provided with written materials that included company history and information explaining the BSC approach and reasons for using it. In all cases, performance was defined completely by measures that were common to both business units across all BSC categories. Using all common measures for divisions with different products and strategies is a "simplifying assumption" used to increase assurance that the evaluation results in this study were determined by the division performance and evaluation focus rather than by the cognitive challenge of working with unique and common measures (Lipe & Salterio, 2000).³

After the participants reviewed the common company scorecard, the definitions for each measure, the specific strategic objectives for each division, and each division's actual results in comparison to target, they rated each division or division president using a scale from 0 ("Dismal") to 100 ("Excellent") – a scale similar to the one used by Lipe and Salterio (2000). The presentation order of divisions was randomized to minimize order effects and no statistically significant ordering effect was found.

After assessing the performance of the evaluation target (i.e., either the division or the president), each participant returned his or her case booklet to the experiment proctor and received a post-experiment questionnaire. In the questionnaire, the participant first rated each performance measure according to a 7-point scale for the level of controllability by the division or division president. This rating helps validate the manipulation of evaluation focus as it pertains to perceived controllability of the measure by the evaluation target.

After rating each performance measure, participants answered various questions relating to their comprehension of the case information and

received monetary rewards based on the number of correct answers. The instructions for the case stated that the participants had the potential to earn cash rewards for putting in good effort and that they would be asked some post-experiment comprehension questions regarding their understanding of the case materials. The purpose of these comprehension questions was to permit further validation checks, to motivate participants to study the case materials, and to eliminate any participants who may not have adequately understood the evaluation task or case context (discussed in the next section). After eliminating the responses of participants with lower comprehension scores, responses on the multiple-choice and scaled judgment questions indicated reasonable comprehension of the case materials and effective manipulation of the target focus variable.⁴ The post-experiment questionnaire also requested demographic characteristics and comments from the participants. The large majority of free response statements gathered in the post-experiment questionnaire provided further indication that participants generally understood the context of the case and responded to manipulations of the target focus.

Participants

A pretest of the case experiment was conducted using 91 Master of Accountancy students. Using a similar post-experiment questionnaire, the pretest evaluation results and responses to comprehension questions regarding the key manipulations in the study led to improvements in the final case materials. Improvements included more specific instructions on what they would be doing in the experiment and altered performance outcomes that included both favorable and unfavorable actual results compared to target (discussed later). More emphasis was given to the focus of the evaluation and additional comprehension questions were added regarding target focus and emphasis on nonfinancial measures in the case. In addition, there were minor changes to make the scorecards more clear.

The actual experiment was conducted during three sections of an MBA course on managerial accounting at a nationally ranked MBA school. Of the 117 students who began the experiment, 110 participants completed both parts (evaluation and post-experiment questionnaire). To help ensure that the results are based on participants who understood their evaluation task, 24 respondents who incorrectly answered the comprehension question on target focus were eliminated. In addition, 14 other respondents with fewer than four correct answers overall on the seven multiple-choice comprehension

questions were eliminated, leaving 72 “informed” respondents. Analysis of the full set of 110 participants demonstrates results that are not qualitatively different from the reduced set of 72 participants, just quantitatively weaker. Thus, the results reported hereafter are those of the “more informed” reduced set of participants.

The average age of the 72 participants was 28.6 years, with an average of 4.5 years of professional work experience and 3.2 years of experience evaluating employees or organizations. Of the 72 participants, 40 (56 percent) had worked in general management-, personnel-, marketing-, or engineering-related fields, and 21 of the 72 (29 percent) had worked in accounting-, finance-, or banking-related industries. Eighty-four percent of participants were male.

Research Design

Each of the 72 participants was randomly assigned an evaluation focus of either the two division presidents or the two divisions. Of the 72 participants, 38 evaluated the division presidents and 34 evaluated the divisions, allowing an analysis of the impact of evaluation focus on the evaluation scores. The president and division evaluation scores served as the primary dependent variables in hypothesis testing.

Table 1 provides the company BSC and example performance results for one of the divisions used in the case. As shown, the scorecard measures used in this case are grouped into FIN measures (net sales growth, operating margin, return on assets, and net cash flow) and nonfinancial measures. The nonfinancial measures are further classified into customer, process, and human resource categories (consistent with the scorecard for the actual company on which the Work-Right case was based). The nonfinancial measures are also segmented into two manipulation groups. Half of the nonfinancial measures (customer satisfaction, on-time delivery, quality index, and employee satisfaction) were manipulated as favorable or unfavorable by setting performance at least 15 percent higher or lower than target, respectively. The other half of the nonfinancial measures (time with customers, manufacturing cycle efficiency, warranty-free sales, and employee suggestions implemented) were not manipulated; and, the actual results were very close to the target results. The purpose of having the nonmanipulated nonfinancial (NNF) measures was to ensure that the number of manipulated nonfinancial (MNF) measures would equal the number of FIN measures.⁵ Hence, all FIN measures were then manipulated similar to MNF measures. Thus, the study employed a $2 \times 2 \times 2$ experimental design using

Table 1. The Balanced Scorecard for the Excavation Division Used in the Experiment.

Measures	2002 Target	2002 Actual	Actual Percent of Target
<i>Financial</i>			
Net sales growth	10.0%	11.5%	115.0
Operating margin	13.0%	15.1%	116.2
Return on assets	15.0%	17.4%	116.0
Net cash flow (\$ million)	\$750	\$900	120.0
<i>Non-Financial</i>			
Customer			
Customer satisfaction survey (MNF)	85.0%	72.1%	84.8
Time spent with customers (hours) (NNF)	1,000	1,006	100.6
Process			
Manufacturing cycle efficiency (NNF)	30.0%	29.8%	99.4
On-time delivery (MNF)	82.0	67.2	82.0
Warranty-free sales (NNF)	90.0	90.4	100.4
Quality index (MNF)	105.0	88.2	84.0
Human Resource			
Employee satisfaction survey (MNF)	1.00	0.82	82.0
Employee suggestions implemented (NNF)	1.50	1.49	99.6

Note: A similar scorecard was prepared for the Commercial Division. MNF, manipulated nonfinancial measures; NNF, nonmanipulated nonfinancial measures. These labels were not used in the actual case materials.

within-subject factors FIN and MNF manipulated as favorable or unfavorable across a dichotomous between-subject factor FOCUS (evaluation focus on either the division or the individual president).⁶

Similar to [Lipe and Salterio's \(2000\)](#) study, the Work-Right case provided participants with measurement data for each division (see appendix for actual and target data and percentages). Target percentages were also balanced between the two divisions so that half of the measures had targets that were 3 percent higher overall for one division while the other half of the target measures were 3 percent higher overall for the other division. In previous BSC studies ([Lipe & Salterio, 2000](#); [Banker et al., 2004](#)) all the measures had actual data above target (i.e., all favorable results), though both studies randomly manipulated the extent of measurement performance results above target. While this study design was similar in some respects to these two earlier studies, to increase the realism of the evaluation task it incorporated actual performance outcomes that were either unfavorable to

target (81.5–85.0 percent of target), close to target (99.4–100.6 percent of target), or favorable to target (115–120 percent of target). The sum of excess performance (percentage above target) for measures within measurement groups was 67.2 percent for the favorable condition and –67.2 percent for the negative condition. Also similar to [Lipe and Salterio's \(2000\)](#) study, the performance outcomes were opposite for the two divisions. For example, if the FIN measurement group was favorable for Excavation Division, then it was unfavorable for Commercial Division (see appendix).

Subsequent analysis of controllability ratings, reported in [Table 2](#), confirm that the participants perceived the MNF measures to be more controllable than the FIN measures. Further, evaluators of divisions reported relatively stronger controllability of nonfinancial measures compared to evaluators of presidents. However, there was no difference in their perceived controllability of FIN measures, which were rated less controllable than nonfinancial measures by both focus groups. These findings are consistent with [Ghosh \(2005\)](#), although Ghosh only considered individual store manager performance (not the entire store's performance). Apparently both groups perceive the FIN measures to have the same lower controllability. Hence, the participants clearly perceived different levels of controllability between financial and nonfinancial measures, and between evaluations focused on divisions versus presidents for the nonfinancial measures.

RESULTS

The Impact of Target Focus on the Weightings of Nonfinancial and Financial Measures

H₁ predicts that performance evaluations of division presidents will be affected more by nonfinancial measures than by FIN measures. H₂ predicts that evaluations of division performance as an operating unit will be affected more by FIN measures than by nonfinancial measures. [Table 3](#) reports the impact of evaluation target focus on least-squares mean evaluation scores for the two types of measures (MNF and FIN) across the evaluation focus (president versus division).

As shown in [Table 3](#), subjects evaluating division presidents were influenced by the MNF measures more than by the FIN measures. The mean overall difference for the MNF measures between the favored and unfavored conditions is statistically significant (difference = 21.85, $p < .001$). In comparison, evaluators of the president's performance were influenced by

Table 2. Mean (Standard Deviation) Controllability Ratings for Scorecard Measures.

	Assessed Controllability Ratings		
	Evaluators of divisions ($n = 68$)	Evaluators of presidents ($n = 76$)	One-tailed p -value
<i>Financial (FIN) Measures</i>			
Net sales growth	4.44 (1.41)	4.54 (1.62)	0.34
Operating margin	5.08 (1.42)	5.00 (1.48)	0.36
Return on assets	4.59 (1.34)	4.65 (1.35)	0.40
Net cash flow	4.43 (1.38)	4.45 (1.54)	0.47
<i>Manipulated Nonfinancial (MNF) Measures</i>			
Customer satisfaction survey	5.77 (1.12)	5.31 (1.36)	0.01
On-time delivery	6.13 (0.86)	5.39 (1.27)	0.00
Quality index	6.10 (0.93)	5.37 (1.29)	0.00
Employee satisfaction survey	5.88 (1.15)	5.69 (1.17)	0.17
<i>Nonmanipulated Nonfinancial (NNF) Measures</i>			
Time spent with customers	6.25 (1.21)	5.42 (1.72)	0.00
Manufacturing cycle efficiency	5.76 (1.24)	5.30 (1.45)	0.02
Warranty-free sales	5.27 (1.62)	4.89 (1.63)	0.08
Employee suggestions implemented	5.94 (1.31)	5.98 (1.24)	0.42
All FIN measures	4.65 (1.21)	4.66 (1.32)	0.49
All MNF measures	5.97 (0.74)	5.47 (1.00)	0.00
p -value	0.00	0.00	

Note: Participants were asked to evaluate each measure in terms of its level of controllability by the division or president (depending on the target focus they were assigned). Means are based on a 7-point Likert scale where 1 = weak and 7 = strong. The FOCUS factor, a between-subject manipulation, required each participant to focus his or her evaluation on either the president of both divisions or on the two divisions overall. Hence, 34 participants evaluated the performance of two divisions, resulting in $n = 68$ for controllability ratings for division evaluators; and 38 participants evaluated the performance of two presidents, resulting in $n = 76$ for controllability ratings for president evaluators.

Paired sample t -tests (one-tailed) were used to compute p -values for statistical differences across the mean score for groups of measures (i.e., all FIN measures to all MNF measures). Similarly, independent samples t -tests (one-tailed) were used to compute p -values for statistical differences across division and president evaluators. Independent sample t -tests were also used to compute p -values across all FIN measures and all MNF measures for division and president evaluators.

There are no statistically significant differences between president evaluators' and division evaluators' assessed controllability of the two employee nonfinancial measures (employee satisfaction survey and employee suggestions implemented). One possible explanation for this "inconsistent consistency" may be indicated by the fact that these employee-based measures are rated as the most controllable of all the measures rated by the evaluators of presidents. Hence, it appears that while division evaluators, compared to the president evaluators, rate nonfinancial measures as more controllable, evaluators feel strongly that presidents can (and should?) exert significant control over employee-based measures.

Table 3. Descriptive Results: The Effects of Evaluation Focus on Evaluation Scores – Impact of Focus on Least-Squares Mean (Standard Error) Evaluation Scores as Performance Measures Vary between Favorable and Unfavorable.

Evaluation Focus	Group of Measures	Mean Score: Favorable Division	Mean Score: Unfavorable Division	Difference	One-Tailed <i>p</i> -Value
President (<i>n</i> = 38)	Nonfinancial (MNF)	77.81 (2.85)	55.96 (2.85)	21.85 (3.90)	< .001
	Financial (FIN)	73.29 (2.85)	60.49 (2.85)	12.81 (3.90)	.001
	Difference between measures (1)	4.52 (2.62)	-4.52 (2.62)	9.04 (5.25)	.045
Division (<i>n</i> = 34)	Nonfinancial (MNF)	74.53 (3.05)	58.62 (3.05)	15.91 (4.18)	< .001
	Financial (FIN)	71.62 (3.06)	61.53 (3.06)	10.09 (4.18)	.009
	Difference between measures (2)	2.91 (3.08)	-2.91 (3.08)	5.81 (6.16)	.825
	Difference between focus (1 -2)			3.23 (8.08)	.345

Note: Participants were asked to evaluate each division or president (depending on the target focus they were assigned). LS means are based on a scale from 0 (“Dismal”) to 100 (“Excellent”). The LS means difference for the Division Focus group (2) is in the opposite direction predicted by the H₂ hypothesis. Hence, the reported *p*-value is the complement of the one-tailed *p*-value test of 5.81 (6.16) difference, or .825 (1-.175).

the FIN measures less strongly, although the difference for FIN measures on evaluation scores was also statistically significant (difference = 12.81, *p* = .001). To test H₁ directly, the mean difference for the MNF measures was compared to the difference for the FIN measures (21.85 versus 12.81, respectively) on evaluations of the division presidents. The 9.04 difference is statistically significant (*p* = .045), which supports H₁. The interaction between the favorable versus unfavorable outcomes and the MNF and FIN measures for evaluations focused on the presidents’ personal performance is illustrated in Fig. 1, Panel A.

As shown in Table 3, similar to those focusing on evaluation of presidents, the influence of both MNF and FIN measures on experiment participants evaluating the divisions was also statistically significant (*p* < .001 and *p* = .009, respectively). However, the influence of MNF versus FIN measures on division evaluations is similar, not dissimilar, to president evaluations. Again, the mean difference for MNF measures is higher than the mean

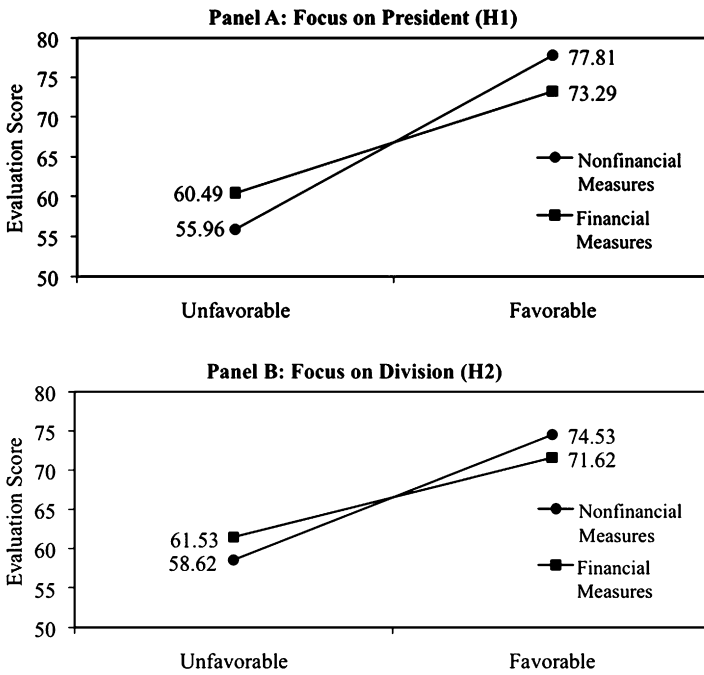


Fig. 1. The Effects of Evaluation Focus on Evaluation Scores.

difference for FIN measures (15.91 versus 10.09, respectively). Clearly, then, H₂ is not supported by these results. Evaluators of divisions do not appear to differentiate between MNF and FIN measures as demonstrated by the analysis of the mean impact of the MNF measures compared to the impact of the FIN measures (difference = 5.81, $p = .825$).⁷ The insignificant interaction between the favorable versus unfavorable outcomes and the MNF and FIN measures for evaluations focused on the divisions' performance is illustrated in Fig. 1, Panel B. The distinction between nonfinancial and FIN measures by those evaluating the overall performance of divisions is weaker than by those evaluating the individual performance of presidents.

DISCUSSION

The purpose of this study is to test whether financial and nonfinancial measures affect BSC performance evaluations differently when the focus of

the evaluation is the individual division president versus when the focus is on the overall division. H_1 proposed that the division president evaluations would be influenced more by nonfinancial measures than by FIN measures. H_2 proposed the opposite – division evaluations would be influenced more by FIN measures than by nonfinancial measures. The results of this study clearly support H_1 ; however, the data do not support H_2 .

The findings of the present study extend the results of Schiff and Hoffman to a BSC context. In a multidivisional BSC performance evaluation, nonfinancial measures clearly affect evaluations that focus on individual performance more than FIN measures. This result is consistent with Schiff and Hoffman (1996), which present evidence that nonfinancial measures are relied on more in personnel-focused evaluations than are FIN measures. However, the findings also suggest that when the focus of the evaluation is on the division, the influence of nonfinancial and FIN measures is not differentiated. Both types of measures appear to influence division evaluators' ratings, but the difference between them is not statistically significant. This result is counter to Schiff and Hoffman's findings.

Schiff and Hoffman link their study to AT. One causal aspect of AT is controllability. Prior research has found that nonfinancial measures are often perceived as more controllable than FIN measures because the FIN measures are too aggregated and the nonfinancial measures are regarded as better signals of the actions managers take to improve overall performance (Ghosh, 2005). In this study, the fact that both groups of evaluators deemed the nonfinancial measures to be more controllable than the FIN measures may help explain why both groups appeared to rely strongly on nonfinancial measures, and in the case of personnel evaluations even emphasize nonfinancial measures over FIN measures. A subsequent analysis, not presented in the body of the chapter, demonstrates that the results do not support an exclusive relationship between the measures of controllability perception and the evaluation scores.⁸ Future studies should consider AT and, in particular, controllability in performance evaluation within the BSC.

The findings also provide implications for companies with multiple divisions using BSCs to evaluate performance. Ittner and Larcker (1998, p. 228) state that an "important question is whether the same measures or scorecard used to develop strategic priorities and monitor strategic actions should be used to evaluate managerial performance." Companies may need to use different scorecards to evaluate manager performance versus evaluating organization performance for resource allocation purposes. When evaluating manager performance, evaluators appear to place more emphasis on measures they deem to be controllable by the manager.

Measures (either financial or nonfinancial) that are perceived to be more controllable by the manager or the organization may (and likely should) have more impact on performance evaluations in comparison to other performance measures.

Limitations

Some limitations of this study should be noted. First, experimental studies such as ours may lack realism and external validity (Runkel & McGrath, 1972) because the actual political setting and reward structures in which evaluators operate are generally missing. However, experimental studies can reduce risks to internal validity by more precisely manipulating the proposed independent variables. Another limitation is that this study focuses on the perspective of the evaluator rather than on that of the performer. Certainly the behavioral impact on the individuals being evaluated is a major issue in BSC theory and should be explored in future research.

Professional interest in BSC management models is widespread, though there is currently limited academic research available to users interested in better understanding the impact of BSC on decision processes and individual and institutional performance. It seems clear, therefore, that more research is required to better understand the impact of BSC in real-world settings. Nevertheless, there are special difficulties involved in conducting such research, including the risks inherent in applying theory to real-world contexts, the need for normative criteria against which evaluations and decisions can be evaluated, necessary scope limitations, lack of proprietary firm data, and the challenge of identifying optimal research participants who represent actual BSC users. This study, as with other studies, is challenged by these difficulties and has, by necessity, required a number of simplifying assumptions in establishing the experimental setting. Lipe and Salterio (2000, pp. 295–296) provided an excellent discussion of research obstacles that must be addressed by future research studies similar to this study.

NOTES

1. Part of the purpose of using two divisions in this study was to investigate the impact of strategic linkage of performance measures. This particular investigation is related to the work of Banker et al. (2004) who found that evaluators respond to more unique measures that are strategically linked. The case in this study was

designed so that nonfinancial measures were linked much more closely to the strategy of the Excavation Division than to the strategy of Commercial Division. Subsequent analysis (unreported in this chapter) finds that the nonfinancial measures had more impact when the strategic linkage was high. Further, the FIN measures also had more impact when the strategic linkage was high. While there is some interest in these data suggesting that a strong strategic linkage for nonfinancial measures adds to the perceived value of the FIN measures, the results are largely a replication of the Banker et al. study and are not the emphasis of this report. Hence, the linkage manipulation is collapsed within the research design in order to concentrate the analysis and discussion on the impact of controllability of performance measures on BSC evaluations.

2. An agreement with the company prevents the disclosure of its name; however, it is a large global company (Fortune 200) with headquarters in the United States. The scorecard used in the experiment is altered somewhat from the original company scorecard to facilitate experiment manipulations.

3. Although BSC theory suggests that strategic business units should have unique BSC measures, business units within a company often end up with the same or similar scorecards. For instance, Malina and Selto (2001) described the top-down implementation of the same scorecard among 31 North American distributorships. In another example, the strategic business units in the Mobil USM&R (A) BSC case (Kaplan, 1996) were given freedom to develop their own scorecards, yet the unit scorecards essentially mirrored the group-level scorecard. Although divisions with different products, markets, distribution channels, or strategies generally do not have the same scorecards, this type of scenario facilitates the testing of the factors of interest in this study.

4. There were seven multiple-choice questions that tested comprehension of certain aspects of the case materials. The average number of correct answers on the multiple-choice test questions was 5.01 out of 7 (standard deviation 1.52). Respondents were given \$1 for each correct answer, a \$2 bonus for six correct answers, and a \$3 bonus for seven correct answers. Actual payouts ranged from \$1 to \$10, with an average payout of \$6.07. Further, six Likert-type judgment questions were used to further evaluate participants' case comprehension, as well as impact of test manipulation. These unrewarded judgment questions were scaled on a 10-point basis from "Strongly Disagree" to "Strongly Agree." The average response in each question was consistent with case intent and test manipulation.

5. The MNF measures were selected based on a pretest showing those measures tended to get rated as having stronger linkage to Excavation Division and a greater perceived linkage spread between Excavation Division and Commercial Division than did the NNF. This is useful for two reasons. First, it established a high-linkage division and a low-linkage division (see Note 1). Second, it permitted the manipulation of the same number of financial and nonfinancial measures, while recognizing at the same time that the BSC approach generally includes more nonfinancial measures than FIN measures.

6. Since this chapter is not intended to exactly replicate prior related papers, the experimental design is not exactly consistent with prior studies. For example, Dilla and Steinbart's (2005) study included a "division" variable as a within-subjects factor (all subjects evaluated the performance of both divisions) and Lipe and

Salterio's (2000) study included a between-subjects factor to test the impact of common and unique measures. While the attempt was made to remain as consistent as possible with prior related research designs, the specific theory being applied in a BSC setting by this study necessitates some differences from research designs used in prior studies.

7. Given that the results on this test are in the opposite direction posited by H_2 , the effective test statistic for H_2 is $p = .825$ (1-.175).

8. Following the guidance of Baron and Kenny (1986), the possibility was evaluated that the subjects' perception of performance measurement controllability (as self-reported in the post-experiment questionnaire) somehow mediated the impact of the experimental manipulation of control (by establishing each subject's focus of evaluation). Measures of the perception of financial measures' controllability and nonfinancial measures' controllability were created by summing the controllability ratings for the four FIN measures and for the four MNF measures for each division. When including these factors in ANOVA testing, the results did not substantiate an exclusive relationship between the two measures of controllability perception and the evaluation scores for the presidents or divisions. Thus, the impact of subjects' perception of controllability does not appear to be distinguished from the experimental manipulation of control via evaluation focus, nor do the subjects' perceptions of performance measurement controllability mediate the relationship between evaluation focus and evaluation score.

ACKNOWLEDGMENTS

We wish to thank two anonymous reviewers, the Associate Editor, Editor Vicky Arnold, Steven Kaplan, Joanna Ho, Mike Petersen, Janel Bloch, Francisco Roman, Nathan Stuart, and Wim Van der Stede; plus seminar participants at Brigham Young University, University of Arkansas, Miami University, Washington State University, Louisiana State University, the Management Accounting Section Research Conference, the Accounting, Behavior, and Organization Section Conference, and the American Accounting Association Annual Meeting for helpful comments on earlier drafts of this chapter.

REFERENCES

- Antle, R., & Demski, J. S. (1988). The controllability principle in responsibility accounting. *The Accounting Review*, 63(4), 700-718.
- Banker, R. D., Chang, H., & Pizzini, M. J. (2004). The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review*, 79(1), 1-23.

- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Brewer, P. (2002). Putting strategy into the balanced scorecard. *Strategic Finance*, January, pp. 44–52.
- Bryant, L., Jones, D. A., & Widener, S. K. (2004). Managing value creation within the firm: An examination of multiple performance measures. *Journal of Management Accounting Research*, 16, 107–131.
- Chow, C. W., & Van Der Stede, W. (2006). The use and usefulness of nonfinancial performance measures. *Management Accounting Quarterly*, 7(3), 1–8.
- DeNisi, A. S. (1996). *Cognitive approach to performance appraisal*. New York, NY: Routledge.
- Dilla, W. N., & Steinbart, P. J. (2005). Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. *Behavioral Research in Accounting*, 17, 43–53.
- Ghosh, D. (2005). Alternative measures of managers' performance, controllability, and outcome effect. *Behavioral Research in Accounting*, 17, 55–70.
- Gumbus, A., & Lyons, B. (2002). The balanced scorecard at Phillips Electronics. *Strategic Finance*, November, pp. 45–49.
- Hopwood, A. G. (1972). An empirical study of the role of accounting data in performance evaluation. *Journal of Accounting Research*, 20(Supplement), 156–182.
- Ittner, C. D., & Larcker, D. F. (1998). Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research*, 10, 205–238.
- Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, 78(3), 725–758.
- Ittner, C. D., Larcker, D. F., & Rajan, M. V. (1997). The choice of performance measures in annual bonus contracts. *The Accounting Review*, 72(2), 231–256.
- Kaplan, R. S. (1996). Mobil USM&R (A): Linking the balanced scorecard. *Harvard Business School Case* (197–025). Boston, MA: Harvard Business School.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard – measures that drive performance. *Harvard Business Review*, January–February, pp. 71–79.
- Kaplan, R. S., & Norton, D. P. (1993). Putting the balanced scorecard to work. *Harvard Business Review*, September–October, pp. 134–147.
- Kaplan, R. S., & Norton, D. P. (1996). *Translating strategy into action: The balanced scorecard*. Boston, MA: Harvard Business School Press.
- Kaplan, R. S., & Norton, D. P. (2001). *The strategy-focused organization*. Boston, MA: Harvard Business School Press.
- Kaplan, S. E., & Reckers, P. M. (1985). An examination of auditor performance evaluation. *The Accounting Review*, 60(3), 477–487.
- Kaplan, S. E., & Reckers, P. M. (1991). An attributional analysis of the performance evaluation process. *Advances in Accounting*, 9, 227–248.
- Kaplan, S. E., Reckers, P. M., & Roark, S. J. (1988). An attribution theory analysis of tax evasion related judgments. *Accounting, Organizations, and Society*, 13(4), 371–379.
- Kent, R. L., & Martinko, M. J. (1995). The measurement of attributions in organizational research. In: M. Martinko (Ed.), *Attribution theory: An organizational perspective* (pp. 7–14). Delray Beach, FL: St. Lucie Press.

- Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review*, 79(4), 1075–1094.
- Lipe, M. G., & Salterio, S. E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review*, 75(3), 283–298.
- Malina, M., & Selto, F. (2001). Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research*, 13, 47–90.
- Martinko, M. (1995). The nature and function of attribution theory within the organizational sciences. In: M. Martinko (Ed.), *Attribution theory: An organizational perspective* (pp. 17–34). Delray Beach, FL: St. Lucie Press.
- Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiasing balanced scorecard evaluations. *Behavioral Research in Accounting*, 16, 75–88.
- Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to method*. New York: Holt, Rinehart and Winston.
- Said, A. A., HassabElnaby, H. R., & Wier, B. (2003). An empirical investigation of the performance consequences of nonfinancial measures. *Journal of Management Accounting Research*, 15, 193–223.
- Salterio, S., & Webb, A. (2003). The balanced scorecard: Rhetoric versus research. *CA Magazine*, August, pp. 39–41.
- Schiff, A. D., & Hoffman, L. R. (1996). An exploration of the use of financial and nonfinancial measures of performance by executives in a service organization. *Behavioral Research in Accounting*, 8, 134–153.
- Silk, S. (1998). Automating the balanced scorecard. *Management Accounting*, May, pp. 38–44.
- Tan, H., & Lipe, M. G. (1997). Outcome effects: The impact of decision process and outcome controllability. *Journal of Behavioral Decision Making*, 10, 315–332.
- Todd, F. T., Houston, R. W., & Peters, M. F. (2001). The impact of internal auditor compensation and role on external auditors' planning judgments and decisions. *Contemporary Accounting Research*, 18(2), 257–281.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Bulletin*, 48(9), 957–965.
- Zimmerman, J. L. (1995). *Accounting for decision making and control*. Chicago, IL: Irwin.

APPENDIX. ACTUAL VERSUS PERFORMANCE TARGETS

For the Excavation Division

Measures	Set 1: FIN+MNF+			Set 2: FIN–MNF+			Set 3: FIN+MNF–			Set 4: FIN– MNF–		
	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target
<i>Financial</i>												
Net sales growth	10.0%	11.5%	115.0	10.0%	8.5%	85.0	10.0%	11.5%	115.0	10.0%	8.5%	85.0
Operating margin	13.0%	15.1%	116.2	13.0%	10.6%	81.5	13.0%	15.1%	116.2	13.0%	10.6%	81.5
Return on assets	15.0%	17.4%	116.0	15.0%	12.5%	83.3	15.0%	17.4%	116.0	15.0%	12.5%	83.3
Net cash flow (\$ million)	\$750	\$900	120.0	\$750	\$622	82.9	\$750	\$900	120.0	\$750	\$622	82.9
<i>Customer</i>												
Customer satisfaction survey	85.0%	98.3%	115.7	85.0%	98.3%	115.7	85.0%	72.1%	84.8	85.0%	72.1%	84.8
Time spent with customers	1,000	1,006	100.6	1,000	1,006	100.6	1,000	1,006	100.6	1,000	1,006	100.6
<i>Process</i>												
Cycle time efficiency	30.0%	29.8%	99.4	30.0%	29.8%	99.4	30.0%	29.8%	99.4	30.0%	29.8%	99.4
On-time delivery	82.0	96.4	117.5	82.0	96.4	117.5	82.0	67.2	82.0	82.0	67.2	82.0
Warranty-free sales	90.0	90.4	100.4	90.0	90.4	100.4	90.0	90.4	100.4	90.0	90.4	100.4
Quality index	105.0	122.9	117.0	105.0	122.9	117.0	105.0	88.2	84.0	105.0	88.2	84.0
<i>Human Resource</i>												
Employee satisfaction survey	1.00	1.17	117.0	1.00	1.17	117.0	1.00	0.82	82.0	1.00	0.82	82.0

APPENDIX. (Continued)

For the Excavation Division												
Measures	Set 1: FIN+MNF+			Set 2: FIN-MNF+			Set 3: FIN+MNF-			Set 4: FIN- MNF-		
	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target
Employee suggestions implemented	1.50	1.49	99.6	1.50	1.49	99.6	1.50	1.49	99.6	1.50	1.49	99.6
Total financial			67.2			-67.2			67.2			-67.2
Total nonfinancial			67.2			67.2			-67.2			-67.2
For the Commercial Division												
Measures	Set 1: FIN-MNF-			Set 2: FIN+MNF-			Set 3: FIN-MNF+			Set 4: FIN+MNF+		
	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target	Target	Actual	Actual percent of target
<i>Financial</i>												
Net sales growth	9.7%	8.2%	85.0	9.7%	11.2%	115.0	9.7%	8.2%	85.0	9.7%	11.2%	115.0
Operating margin	13.4%	10.9%	81.5	13.4%	15.6%	116.2	13.4%	10.9%	81.5	13.4%	15.6%	116.2
Return on assets	15.5%	12.9%	83.3	15.5%	17.9%	116.0	15.5%	12.9%	83.3	15.5%	17.9%	116.0
Net cash flow (\$ million)	\$728	\$603	82.9	\$728	\$873	120.0	\$728	\$603	82.9	\$728	\$873	120.0

<i>Customer</i>												
Customer satisfaction survey	82.5%	69.9%	84.8	82.5%	69.9%	84.8	82.5%	95.4%	115.7	82.5%	95.4%	115.7
Time spent with customers	1,030	1,036	100.6	1,030	1,036	100.6	1,030	1,036	100.6	1,030	1,036	100.6
<i>Process</i>												
Cycle time efficiency	30.9%	30.7%	99.4	30.9%	30.7%	99.4	30.9%	30.7%	99.4	30.9%	30.7%	99.4
On-time delivery	79.5	65.2	82.0	79.5	65.2	82.0	79.5	93.5	117.5	79.5	93.5	117.5
Warranty-free sales	87.3	87.6	100.4	87.3	87.6	100.4	87.3	87.6	100.4	87.3	87.6	100.4
Quality index	108.2	90.8	84.0	108.2	90.8	84.0	108.2	126.5	117.0	108.2	126.5	117.0
<i>Human resource</i>												
Employee satisfaction survey	1.03	0.84	82.0	1.03	0.84	82.0	1.03	1.21	117.0	1.03	1.21	117.0
Employee suggestions implemented	1.46	1.45	99.6	1.46	1.45	99.6	1.46	1.45	99.6	1.46	1.45	99.6
Total financial			-67.2			67.2			-67.2			67.2
Total nonfinancial			-67.2			-67.2			67.2			67.2