Jaap Scheerens

Educational Effectiveness and Ineffectiveness

A Critical Review of the Knowledge Base

Springer

Educational Effectiveness and Ineffectiveness

Jaap Scheerens

Educational Effectiveness and Ineffectiveness

A Critical Review of the Knowledge Base

Springer

Jaap Scheerens
Faculty of Behavioural, Management
and Social Sciences
University of Twente
Enschede
The Netherlands

ISBN 978-94-017-7457-4 ISBN 978-94-017-7459-8 (eBook)
DOI 10.1007/978-94-017-7459-8

Library of Congress Control Number: 2015954351

Springer Dordrecht Heidelberg New York London

Printed on acid-free paper

Springer Science+Business Media B.V. Dordrecht is part of Springer Science+Business Media (www.springer.com)

Preface

In a set of reviews on the state of the art of educational effectiveness research in the Journal *School Effectiveness and School Improvement* (June 2014), the field is described as "in good shape" (Reynolds et al. 2014, 196). In my own commentary of the reviews this positive message was partly confirmed: there is strong consensus on the set of malleable conditions associated with good student performance. Yet, I thought that some critical comments were warranted as well: for most of the identified and generally supported malleable factors, quantitative meta-analyses show considerable differences in the estimate of average effect sizes, and the progress in the direction of a more model- and theory-driven approach is quite slow (Scheerens 2014). In this book the challenges presented by the cited review studies are taken up by means of more detailed and explicit modelling and an extensive review of the literature that spans four decades of research (Part I), a presentation and review of quantitative meta-analyses (Part II) and further reflection on theoretical foundations as well as analyses of practical application of the empirical results (Part III). The work is based on research carried out over a period of about 10 years as part of a research program at the University of Twente in the Netherlands, led by the author, comprising several meta-analyses, review studies and secondary analyses of international datasets. In the course of synthesizing this work it became clear that the reality that was emerging was not that of a unilateral success story. Despite "positive results" in the sense of support for effective malleability and the straightforward logic of the conceptual models, considerable "dissonance" was noted as well, in the form of strong variation in research results for most key factors, very low effect sizes resulting from our own studies and an influence of "given" background and contextual conditions that was frequently considerably stronger than that of malleable, policy amenable variables. These results prompted a growing interest in the question of why so many plausible malleable variables did not work in so many instances. And this growing preoccupation with "ineffectiveness" was still enforced by experiences with secondary analyses on international datasets. There we found negligible effects for most of the school characteristics in most countries, with only very few exceptions, lack of

change in average performance over time, in most countries, and failure to identify indirect effects of system-level levers, hypothetically mediated by conditions at school level. Such a reading of the international results in question contrasts with the image of strong positive malleability in some reports by the OECD and McKinsey. These observations stimulated the search for theoretical mechanisms, able to explain not just the positive results and effectiveness but also the disappointing results and "ineffectiveness". When addressing application of the educational effectiveness research findings, the issue of research utilization was encountered as well as the connection between effectiveness research and school improvement. It was noted that recommended approaches for school improvement and systemic reform differ in the degree to which they remain close to the empirical evidence. Some cases that were mentioned appeared to go far beyond the educational research evidence in overstressing the impact of secondary functions in schooling, such as leadership and cooperation between teachers. A case study of policies and bottom-up developments in Dutch education further illustrates the complexities of systemic reform when the ambitions of evidence based work are clashing with strong school autonomy.

The author is indebted to Rien Steen for his contribution to Chap. 8, to Dr. Hans Luyten, Dr. Maria Hendriks and Prof. Dr. Cees Glas who were involved in various parts of the research that was reported in this book and to Prof. Dr. Greetje van der Werf for reviewing a draft of the final chapter. Next the author acknowledges the permission provided by Taylor and Francis http://www.tandfonline.com/, to use material from articles published in the journals *School Effectiveness and School Improvement* and *School Leadership and Management*; full references of which will be provided in the chapters where this material is used.

References

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., Teddlie, C., & Stringfield, S. (2012). Educational effectiveness research (EER): A state of the art review. *School Effectiveness and School Improvement*, *25*, 197–230.

Scheerens, J. (2014). School, teaching, and system effectiveness: some comments on three state-of-the-art reviews. School Effectiveness and School Improvement, *25*(2), 282–290.

Contents

Part II Quantitative Research Results

Part III Theoretical Interpretation and Practical Application

About the Author

Jaap Scheerens is Professor Emeritus at the University of Twente in the Netherlands. After his retirement he was Guest Lecturer at the Universities of Bristol and Rome (Roma Tre). He is currently associated with the research institute Oberon in Utrecht. At the University of Twente he coordinated a research program on school effectiveness. During his career he was scientific director of the national research school for postgraduate training in educational science ICO and director of the research institute OCTO of the faculty of education at the University of Twente. He was involved in many international research projects, funded by international organizations. He has published 20 books and about 100 articles in scientific journals, mainly addressing educational effectiveness and educational evaluation.

Part I
Modelling Educational Effectiveness at Teaching, School and System Level

Introduction to Part I

Educational effectiveness is modelled according to a multi-level framework, comprising four levels: the individual student level, the level of classroom teaching, the school and the national educational system. This overall framework is introduced in Chap. 1. The overall hierarchical structure includes steering and control measures between the levels, and antecedent and ecological context conditions. At each of the levels production processes are modelled according to a context, input, process and output scheme. In the subsequent chapters partial models at the level of teaching and learning (Chaps. 2 and 3), the level of the school (Chaps. 4 and 5), and the level of national educational systems (Chap. 6) are presented. The conceptual models function as ordering frameworks of the variables that have been empirically studied over a period of four decades. Apart from formal modelling aspects, this part of the book consists of review of research and theoretical contributions to teaching, school and system effectiveness. This overview yields various selections of key independent variables at each of the levels; some more theory oriented, and others more plain summaries of the variables most frequently used in research. A final selection of variables that reflect effectiveness enhancing variables at classroom and school level is described in more detail. This set of variables is used in the meta-analyses that are presented in Part II, Chap. 8. The review and description of key variables is an update of earlier published and unpublished work, Scheerens and Bosker (1997, Chap. 4), and Scheerens et al. (2005) respectively.

References

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., Seidel, T., Witziers, B., Hendriks, M., & Doornekamp, G. (2005). *Positioning the supervision frameworks for primary and secondary education of the Dutch Educational Inspectorate in current educational discourse and validating core indicators against the knowledge base of educational effectiveness research*. Enschede/Kiel: University of Twente/Institute for Science Education (IPN).

Chapter 1
An Overarching Conceptual Framework

Keywords Educational effectiveness · Teaching effectiveness · School effectiveness · System effectiveness · Education as a hierarchical system · Autonomy · Alignment · Loose coupling · Planning and control models · Educational ineffectiveness

Introduction

Educational effectiveness is one of the measurable facets of the larger concept of educational quality. This book discusses the knowledge that has been obtained in empirical studies on school and teaching effectiveness. In the first chapter a conceptual framework is presented that allows for a sensible organization of this knowledge. An important new element is the inclusion of "system effectiveness", addressing the question of the effectiveness of educational policies and structures at national level. One of the functions of the conceptual framework is to set the stage for an integration of system, school and teaching effectiveness. The structure of the model is therefore hierarchical, and the approach could be termed "systemic". In the process of explaining this hierarchical structure it becomes clear that systemic should not be seen as referring to a smoothly running engine controlled from the top. More complexity is involved: next to steering impulses from higher levels, autonomy of lower levels and varying degrees of "loose coupling" between levels are key features of the conceptual framework.

Educational Systems as Hierarchical Structures

Educational systems can be seen a hierarchical structures, with loosely connected core processes at each level. Starting from the bottom up, the learning processes of individual students at school could be taken as a starting point. Although learning

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_1

processes are by no means limited to "formal" learning, the emphasis will be on formal learning.[1] In a basic, perhaps abstract way, formal and informal learning are the same in evoking changes in the psychological state of the learner that may be conscious or unconscious, and of which a certain stability and endurance is expected. Formal learning can also be seen as controlled learning, where "control" would be defined in a basic, relatively undemanding way, namely as "any change that can be subjected to an evaluation" (De Leeuw 1990, p. 107); a condition that is already met when any kind of actor, using any kind of procedure, could lead to a judgment about a change from state A to state B.

Under the influence of "constructivism", the image of learning as a constructive process of a learning subject has become generally accepted. The implication is that student learning should basically be seen as a quasi-autonomous process that is self-regulated.

Teaching, situated at the next layer up of our hierarchical system, should be seen as a "booster" of such self-regulated processes. This is another way of expressing the earlier mentioned loose coupling between the hierarchical levels of educational systems, pointing in this case to the relationship between a teaching teacher, and a learning student. The idea of a booster symbolizes the idea that higher level interference is just an external stimulus to an already "running engine". Yet, different schools of thought and teaching models (e.g. the model of direct teaching) stress the importance of structuring learning tasks for students; and actual teaching and learning situations can be placed on a continuum that runs from structure to independence. The learning environment, at the level of classrooms or learning groups is seen as determined by direct "manipulations" of teachers, as well as influenced by the classroom "ecology", which is defined in terms of compositional factors, general atmosphere or climate and interactions between both kinds of factors. Teacher background characteristics, such as their qualifications and experience, have an impact on the learning environment as well.

In traditional models of the school, professional autonomy of teachers stood out as a defining characteristic. But in more recent perspectives, which emphasize distributed leadership at school and "teacher leadership", this relative autonomy is prominently present as well. At the same time schools are currently being seen as managed organizations, in which a central mission, coordination, school evaluation and monitoring, as well as curricular guidelines are meant to facilitate the work of teachers, but might at the same time be perceived as constraining teacher autonomy. So, as was the case for teacher student relationships, the working relationship between school leaders and teachers shows the same dynamic tension between guidance and pre-structuring on the one hand and independent functioning on the

[1]Formal learning at school is taken as the situation in which students are expected to work on specifically designed tasks with the intention of furthering knowledge, skills and personal development.

other. Student composition, climate and ecological factors are distinguished as more holistic and interactive dimensions of the school environment, and they are seen as a separate category, next to direct teaching acts.

At the top level of our four layered system, the level of national or state/regional educational governance, structural reform measures concerning decentralization and accountability and evaluation arrangements have dominated the agenda's during the last two decades. Once again we are confronted with mixed views and mixed patterns regarding higher level control versus providing autonomy to lower layers in the system. This is fruitfully captured in the concept of functional decentralization, which recognizes that systems may centralize in one functional area, for instance the curriculum, and centralize in other functional domains, such as the financial management of schools.

When educational systems are represented as hierarchical, loosely coupled systems, the following types of interrelationships can be highlighted:

- under the premise that education can be abstractly described as a production process, ultimately leading to student outcomes, functioning at each level can be characterized as managed or malleable processes and inputs, leading to outcomes defined at the same object level or a lower level, under the constraint of given antecedents and contextual conditions;
- across level relationships are seen as direct control measures from a higher level to the next down lower level, and as the shaping of already available antecedent conditions into certain "ecologies" (this concept will be explained further on); ideally feedback loops would inform active policies at higher level about outcomes at lower level;
- the idea of loose coupling is based on the assumption that at each level actors have autonomy, so that object level functioning is far from being totally controlled and determined by a higher level, but takes leeway for endogenous development and self-organization;
- an ideal type interpretation of this kind of loose coupling is the idea of subsidiarity, which states that everything that could reasonably be carried out at a lower level should not be taken over by a higher level; schools of thought on educational policy, school management, teaching and learning differ, however, in the way they would define "reasonable" in this definition, and we need to turn to the analytical and empirical literature to get further ideas as to where structure and where independence is expected to be most effective.

The Overall Framework

The elementary design of educational effectiveness research is the association of hypothetical effectiveness-enhancing conditions and output measures, mostly student achievement. The basic model from systems theory, shown in Fig. 1.1, is

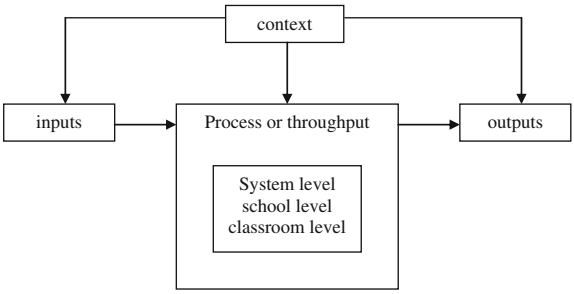

Fig. 1.1 A basic systems model on the functioning of education

helpful in clarifying this design. The major task of educational effectiveness research is to reveal the impact of relevant input characteristics on output and to "break open" the black box in order to show which process or throughput factors "work", next to the impact of contextual conditions. The model, shown in Fig. 1.1, can be used at different levels of aggregation. In the figure this is indicated by mentioning three levels in the central black box of the model: the level of a national educational system, the school level and the level of the instructional setting, often indicated as the classroom level. The three levels are nested, in the sense that schools function within an educational system at national level and classrooms function within schools.

The idea of an educational system as a set of nested layers is made more explicit in variations of the basic model, shown in Fig. 1.1.

The model that was used in the OECD education indicator project (INES), is a case in point (see Fig. 1.2). The model has two dimensions: aggregation level of key actors in the educational system (the national level, the school level, the level of instructional settings and the level of individual participants in education), and aspects of educational productivity, outputs, malleable processes and inputs and

	Outputs	**Malleable inputs and processes**	**Antecedent and contextual conditions**
The national educational system	e.g. national average in reading performance	e.g. evaluative potential at national level	e.g. cultural homogeneity
Educational institutions	e.g. average reading performance per school	e.g. instructional leadership	e.g. school composition in terms of SES
Instructional settings	e.g. average reading achievement per grade or class	e.g. structured teaching	e.g. classroom composition in terms of SES
Individual participants	e.g. reading achievement	e.g. time on task in reading lessons	e.g. cultural capital

Fig. 1.2 The INES model (adapted from Luyten et al. OECD 2005)

antecedent or contextual conditions. The advantage of this model formulation is that it recognizes the fact that input, process and outcome indicators can be described at different levels, including the individual student level.

The structure of this model is fit to define core policy issues in education, which can also be seen as measurable facets of educational *quality*:

- *educational productivity* can be highlighted by focusing on output variables at different aggregation levels, the well-known comparisons between mean performance levels in educational assessment studies between countries, such as TIMSS and PISA, are examples of comparisons of national systems on educational productivity;
- *educational effectiveness* would seek to determine the "net" effect of malleable educational conditions on outputs, while controlling for relevant antecedent conditions at the level of individual participants;
- *educational equity* is captured by examining disparities between resources and processes as well as the variation between students and schools in educational outputs; and the degree to which achievement levels and disparities hang together with specific antecedents of students, schools and school contexts; e.g. the reading performance of girls from cultural minority background, the average achievement levels of schools in rural areas;
- *educational efficiency* would address questions of input provision and effectiveness at the lowest possible costs.

The structure can be used as a framework to generate more specific issues of educational policy as well. The third column of policy malleable conditions at different levels is particularly useful for this. Examples of policy issues at different levels of the framework are: policies of decentralization and accountability at national level, educational leadership and student selectivity at school level, teaching of learning to learn strategies at the level of instructional settings, and engagement at the level of individual students. At the same time policy issues may be determined by the interest in specific outputs, e.g. attitudes towards reading, or interest in specific subgroups of students (e.g. girls, cultural minority students).

Alternative Representation and Expansion of the Basic Framework

The conceptual framework is further elaborated by examining its overall structure, bearing in mind the most relevant empirical research evidence and, finally, by analysing theoretical positions and explanatory mechanisms.

A discussion on the dynamic aspects of the conceptual framework, as shown in Fig. 1.1, is facilitated by making the across level influences more visible. This is done in Fig. 1.3. Influence across levels is indicated by the dotted arrows that run

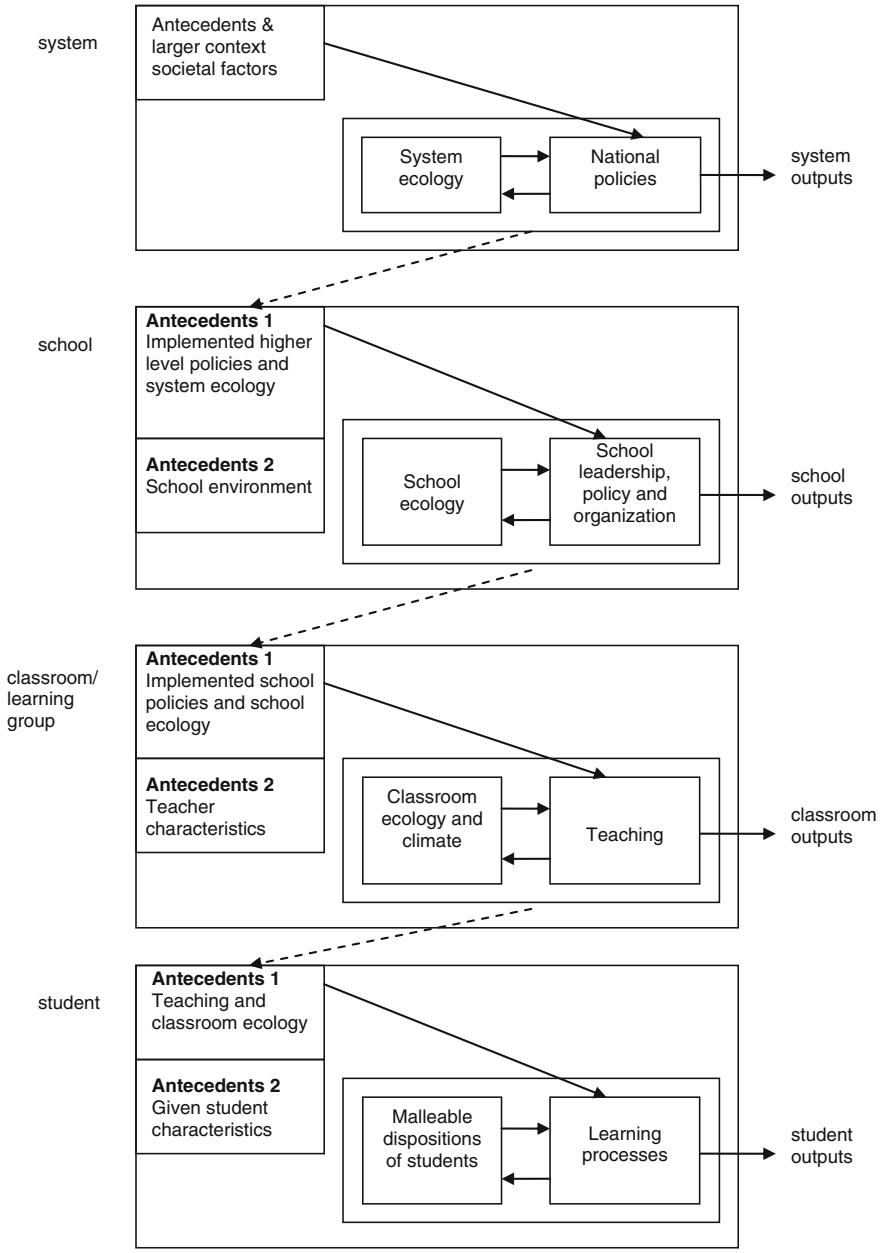

Fig. 1.3 Integrated multi-level model of education; *solid arrows* represent managed control actions, the *dotted arrows* from one system level to the next represent across level influences; feedback-loops (not shown in the diagram) are assumed to run from outcomes at each level to the box containing ecology and active policies at each object level and from lower to higher level

from higher levels to lower levels. Such across level relationships can be interpreted in terms of control, facilitation and buffering from a higher level directed at the core process at the next lower level.

Depicting education in this way emphasizes that the overall system is seen as hierarchical: higher levels impact on lower levels, which, in their turn, may be influenced by what happens at lower levels, through feedback loops. Despite the notion of higher level control, lower levels are seen as having considerable discretion over their core processes, in other words considerable autonomy, which can also be expressed as loose coupling between hierarchical levels.

The degree of higher level control versus lower level autonomy is an issue of central importance at all levels. At system level it is about effective patterns of functional decentralization, which means that, perhaps dependent on the larger context, certain patterns of centralization in some functional domain (e.g. the curriculum) and decentralization in another domain (e.g. financial management) work best. At school level it is about the degree of participative decision-making, or "distributed leadership", and at classroom level it refers to the balance between strongly structured didactic approaches and more open teaching and learning situations that are expected to invite self-regulated learning. Structure versus independence is a red line that dominates policy and research agendas in education.

A second key element in the representation in Fig. 1.3 is the identification of ecology as a separate class of conditions influencing educational performance. This is done by giving a more explicit place to partially controllable composition effects, and their interaction with malleable variables, such as the school climate. The recognition of this kind of contextual conditions emphasizes the partiality of direct control in education, and in this way underlines the loose coupling between the hierarchical levels, but at the same time focuses the attention on a qualitatively different strand of control measures, namely those of selection, admission, grouping and matching of teachers and subgroups of students, as well as on cultural aspects associated with student and teacher body composition. The strength of the interaction of composition and malleable variables, for example in the form of the joint impact of a "good" relational climate at school and a high average socio-economic status of the school or classroom on student performance, has an equity relevant interpretation as well. Interestingly, as a thematic study on PISA 2000 showed, in some countries the interaction between average SES and "good" teaching conditions is much stronger than in other countries; implying that these latter countries do better in creating equitable conditions in education (Luyten et al. 2005).

To underline this point, context and ecology is proposed as an additional column in the basic framework that was depicted in Fig. 1.2. This comes down to separating antecedent conditions and contextual conditions, which, in Fig. 1.2, are united in one column. This is shown in Fig. 1.4.

	Outputs	Malleable inputs and processes	Antecedent conditions	Ecology and contextual conditions
Individual participants	e.g. reading achievement	e.g. time on task in reading lessons	e.g. scholastic aptitude	e.g. cultural Capital at home
Instructional settings	e.g. average reading achievement per grade or class	e.g. structured teaching	e.g. classroom composition in terms of SES	e.g. interaction of classroom composition and classroom climate
Educational institutions	e.g. average reading performance per school	e.g. instructional leadership	e.g. school composition in terms of SES	e.g. interaction of school composition (SES) and school climate
The national educational system	e.g. national average in reading performance	e.g. evaluative potential at national level	e.g. cultural homogeneity of the population	The way education is valued in the society.

Fig. 1.4 Expanded version of Fig. 1.2

A Closer Look at Effectiveness and Efficiency Interpretations of the Framework

The framework described in the previous sections was indicated as covering *educational quality* in a broad sense, and various measurable interpretations (productivity, effectiveness, efficiency and equity) were mentioned. As the core theme of this book is educational effectiveness, this perspective will be further clarified. In addition it will be compared to the, closely associated, perspective of cost-effectiveness and efficiency in education.

The Effectiveness Perspective: Specification of Input-Process-Context-Output Models at Student, Classroom, School and System Level

According to the framework each separate "object level" is characterized by antecedent, contextual and malleable conditions, which are expected to affect student performance. Different strands of educational effectiveness research concentrate on the impact of malleable inputs and processes, on student performance, taking into consideration antecedent and conditional conditions. These different strands are, respectively, instructional effectiveness research and research on teaching, school effectiveness research and research into the effectiveness of

macro-level conditions of educational systems, also to be indicated as "system effectiveness". Instructional effectiveness is to be seen in close connection to conceptualizations of student learning, and effective learning strategies. Research studies within these different strands provide a basis for selecting the most relevant malleable variables, and provide insights in the importance of certain antecedent and contextual conditions as well. Theoretical contributions highlight explanatory mechanisms and more synthetic descriptions of key issues.

General Definition

In order to explain the effectiveness perspective further, school effectiveness will be addressed in more detail.

In the most general sense "school effectiveness" refers to the level of goal attainment of a school. Although average achievement scores in core subjects, established at the end of a fixed programme are the most probable "school effects", alternative criteria like the responsiveness of the school to the community and the satisfaction of the teachers may also be considered.

Assessment of school effects occurs in various types of applied contexts, like the evaluation of school improvement programmes or comparing schools for accountability purposes, by governments, municipalities or individual schools.

School effectiveness *research* attempts to deal with the causal aspects inherent in the effectiveness concept by means of scientific methods. Not only is assessment of school effects considered, but particularly the attribution of differences in school effects to malleable conditions. Usually, school effects are assessed in a comparative way, e.g. by comparing average achievement scores between schools. In order to determine the "net" effect of malleable conditions, like the use of different teaching methods or a particular form of school management, achievement measures have to be adjusted for intake differences between schools. For this purpose student background characteristics like socio-economic status, general scholastic aptitude or initial achievement in a subject are used as control variables. This type of statistical adjustment in research studies has an applied parallel in striving for "fair comparisons" between schools, known under the label of "*value added*".

Demarcation Between School Effectiveness and Educational Effectiveness

When educational systems are seen as hierarchies, school effectiveness can be distinguished from instructional effectiveness, which plays out at classroom level, and from "system effectiveness". The latter term is less common, and refers to a more recent strand of research that is strongly stimulated by the upsurge of international assessment studies. In such studies policy amenable conditions at the national system level can be associated with student outcomes; examples are

policies of enhancing school autonomy, accountability and choice. When school effectiveness depends on school-level malleable conditions, instructional (or teaching) effectiveness on activities of teachers, and system effectiveness on policy amenable conditions at the national level, the term *educational effectiveness* can be used as referring to the union of these three.

As far as research methodology is concerned, multi-level analysis has contributed significantly to the development of integrated school effectiveness models. In contributions to the conceptual modelling of school effectiveness, schools became depicted as a set of "nested layers" (Purkey and Smith 1983), where the central assumption is that higher organizational levels facilitate effectiveness enhancing conditions at lower levels (Scheerens and Creemers 1989). From the overarching perspective of educational effectiveness, it is considered more interesting and policy relevant to see school-level factors in relation to system level and classroom level variables. This approach could either be described as confirming to the conceptual modelling of integrated school effectiveness models, or as treating school effectiveness as embedded in educational effectiveness.

Demarcation Between School Effectiveness and School Improvement

The concept of school improvement may refer to a *product* (improved performance of a school over time), or to a controlled or emerging *process of change* that evolves in time, involving procedural aspects and specific content.

When school effectiveness is seen as a research activity; school improvement could be taken as the dynamic application of the research results, i.e. the active manipulation of the "process" conditions identified as correlates of educational outcomes. A first and basic view of linking improvement and effectiveness would therefore be to say that the results of school effectiveness research provide likely content for school improvement. When school improvement is seen as a systematic activity, two extra emphases are usually at stake; first that the process of setting in motion effectiveness-enhancing conditions is studied as a change process, and second that the control of the change process is seen as distinct from routine control of the organization. This means that school improvement goes beyond the direct management of the primary process of teaching and learning but often includes adaptations of the management approach and organizational conditions as well.

The growing interest in both fields (educational effectiveness and school improvement) in longitudinal designs, often referred to as a more *dynamic approach,* narrows the distinction between them, and makes a complete conceptual integration more feasible (Creemers and Kyriakides 2012). The role of school management and leadership in school improvement is particularly interesting. In some conceptual models (e.g. Hallinger and Heck 2010), improvement is the result of school leadership efforts, while changes in leadership approach might also be seen as part of a school improvement program. External "change agents" may be involved in the latter case.

The Efficiency Perspective: Effectiveness at the Lowest Possible Costs

Cost-effectiveness studies in education are designed to assign priorities to those practices or policies, which use available resources most efficiently. To do so one must know not only the effectiveness of different practices, such as teacher characteristics, instructional strategies or managerial autonomy, but also their costs. Resources in education are always limited, and applying them to the most efficient practices means that the overall educational result will be highest for any given resource outlay or budget.

In the simplest sense the additional element in efficiency oriented studies, over and above identifying effective policies and programmes would be to obtain information on actual expenditure or cost estimates of those policies and programmes.

Leaving aside considerable methodological difficulties in both aspects: establishing which conditions of schooling are effective and cost accounting and estimation relative to these, there is also more to the issue in a conceptual sense.

The efficiency question can be formulated in two distinct ways:

- how to improve outcomes, without increasing costs
- how to produce the same level of outcomes at lower costs

The first orientation fits best with the more common approach to educational improvement and the enhancement of school effectiveness in education. Since cost-effectiveness work is usually carried out by economists, it is relevant to note that economics has a theoretical explanation for stimulating effectiveness and efficiency. This is evident, for example in the work on public choice (Niskanen 1971), which states that public sector organizations tend to be inefficient due to the inherent incentives for budget maximization and overproduction of services, while at the same time review processes are considered as deficient. Introduction of market mechanisms is seen as the basic remedy. This is why efficiency oriented work in education would have an interest in a specific set of educational policy measures, namely those that are related to privatization, enhancing school autonomy, accountability, the use of voucher systems and free school choice. The proof of the pudding of these measures should be that they would stimulate effective production, while diminishing costs at the same time, e.g. by avoiding unnecessary administration costs.

From the second line of thinking, where efficiency gains in education are primarily seen in terms of cost reduction, while maintaining the same level of outcomes, there would be a particular fascination with those inputs and processes that are very costly. This explains the strong interest in the effects of class size and teacher salaries.

In summary, economic research related to cost-effectiveness in the public sector implies a research approach that combines establishing effectiveness and accounting for costs. Second, it provides substantive suggestions with respect to the

Table 1.1 Selection of variables from an economist's perspective

Efficiency enhancing variables from the perspective of microeconomic theory	Cost-relevant educational conditions
– Private schooling	– Teacher salaries
– Private tuition	– Pupil teacher ratio, class size
– School autonomy, including discretionary authority of principals over hiring and firing of teachers and aspects of their wages	– School size
– Free school choice of parents	– Teachers' qualification levels
– Vouchers	– ICT provisions
– Accountability provisions, including rewards and sanctions associated with school and teacher performance	– Support staff, managerial overhead costs of schools, administration costs
	– School funding formula related to school composition

selection of malleable variables that are relevant, in the sense of being cost-sensitive and stimulating effectiveness. Examples of such variables are summarized in Table 1.1.

Theoretical Concepts and Interpretations

When it comes to selecting the most relevant variables in the input-process-output models at each aggregation level, selection could be guided by criteria of theoretical relevance and empirical support (e.g. average effect size in meta-analyses). Consideration of empirical support could be done on the basis of the results of available meta-analyses and reviews. Theoretical relevance would depend on the identification of the more basic principles and mechanisms that lay behind the multi-level representation of educational systems.

To set the stage for the identification of such theoretical principles the four object level models (learning, teaching, school management and system level policy) were combined in a multi-level framework, as was shown in Fig. 1.3. The following facets shed more light on the structure and basic mechanisms that could be related to this framework.

Vertical Control Mix

Joining of the four object level models allows for the recognition of vertical across level relationships.

Different interpretations of the across level influences can be given. A first way to address these interpretations is the distinction between goal control, adaptive

control and routine control made in control theory. Assuming that higher levels control lower levels, higher levels are seen as controllers and lower levels as controlled systems. The degree of control that controllers have over their controlled systems depends on the scope and quantity of control measures. If the central government controls the goals, structures and primary processes, schools would just have to carry out guidelines. More realistically central government would determine basic structures, like rules to found and run a school, and admission policies, but leave primary processes largely to the discretion of individual schools, more or less constrained by central curricula. Goals could come in the form of very general normative orientations or as specific standards, and patterns could be different for various subject matter domains and aspired competencies.

Apart from interpreting across level influences as control, they may also be seen in terms of facilitation, technical support and buffering. Facilitation refers to the provision of adequate human resources, as when the state takes care of the quality of teacher training, as well as the provision of financial and material resources. Technical support could come in areas like in service training, school guidance and counselling, and assistance to deal with students that require special care. Buffering should be seen as a higher level protecting a lower level from too much interference from the environment. Examples are school heads taking care of government bureaucracy, and ministries of education filtering the fluctuating demands on education and training from the world of labour.

Direct and Indirect Effects

The model depicted in Fig. 1.3 includes direct and indirect effects. Direct effects address the bivariate association of a particular independent variable (a policy amenable variable or a contextual variable) and a dependent variable. Indirect effects include third variables that are seen as mediating the influence of the independent variable. The dependent variable is a specific measure of student achievement or attainment. A central notion of the model is across level mediation, when for example the effects of educational policy measures at system level are seen as being mediated by school management and instructional conditions. However, indirect effects can also be studied at one particular aggregation level, for example when the collective efficacy of teachers is seen as mediating the effects of school leadership. Indirect effect models "open black boxes", in order to better explain how higher level conditions, or prime malleable conditions influence student outcomes. However, direct effects are very interesting as well. Path analysis techniques and structural equation modelling sometimes produce results that indicate that effects can be both direct and indirect. When direct effects are not completely accounted for by indirect effects, this could be interpreted as a failure of the mediation hypotheses, but also as the presence of a "genuine", at least substantively interpretable impact on the dependent variable.

Degree of Higher Level Pre-structuring

An issue that runs through the presentation of the models at each level is the issue of control versus autonomy or structure versus independence. At national system level this issue presents itself in the form of functional decentralization models, at school level it is about the degree of framing and guidance that is imposed on the traditional professional autonomy of teachers and at classroom level it presents the competition between structured didactic approaches versus the creation of more open learning situations to foster self-regulated learning. The way the multi-level framework was presented, and further specified at the four different levels made it clear from the outset that educational systems cannot be seen as totally controlled, but are more fruitfully interpreted in terms of multiple relatively autonomous layers that are hierarchically arranged. Thereby structure versus independence is to be seen in terms of continua, with different interpretations at each level, and not simplistically as discrete opposites.

Planning models or control principles can be seen as management theories that have a position on the various continua running from structure to independence, Put differently: prescriptive ideas concerning planning and management in education vary in the degree of control that higher levels exercise over lower levels.

Synoptic Planning

The ideal of "synoptic" planning is to conceptualize a broad spectrum of long term goals and possible means to attain these goals. Scientific knowledge about instrumental relationships is thought to play an important role in the selection of alternatives. Campbell's (1969) notion of "reforms as experiments" combines a rational planning approach to social (e.g. educational) innovation with the scientific approach of (quasi-) experimentation.

The main characteristics of synoptic planning as a prescriptive principal conducive to effective (in the sense of productive) organizational functioning, as applied to education, are:

- "proactive" statement of goals, careful deduction of concrete goals, operational objectives and assessment instruments;
- decomposition of subject matter, creating sequences in such a way that intermediate and ultimate objectives are approached systematically;
- alignment of teaching methods (design of didactical situations) to subject matter segments;
- monitoring of the learning progress of students, preferably by means of objective tests.

The synoptic planning approach in education applies most of all to curriculum planning, design of textbooks, instructional design and preparation of (series of) lessons. Innovations are mainly implemented by central impulses, accompanied by

conformity control (inspection). Improvement is ultimately understood in terms of optimizing the primary process of teaching and learning on the basis of a systematic proactive process of planning and design.

When the ideal of rational planning is extended to organizational structuring, related principles about "controlled arrangements" are applied to the division of work, the formation of units and the way supervision is given shape. "Mechanistic structure", "scientific management" and "machine bureaucracy" are the organizational–structural pendants of rational planning (cf. Morgan 1986, Chap. 2). The basic ideas go back to Max Weber, who stated the principles of bureaucracy as "a form of organization that emphasises precision, speed, clarity, regularity, reliability, and efficiency achieved through the creation of a fixed division of tasks, hierarchical supervision, and detailed rules and regulations". Although Mintzberg's (1979) conception of the professional bureaucracy, applicable to schools and universities, is often treated as the complete antithesis of classical bureaucracy, it should be underlined that the basic notion of standardization and predictability of work processes, be it with a considerable bandwidth of individual leeway, is retained.

The Cybernetic Principle

A less demanding type of planning than synoptic planning is the practice of using evaluative information on organizational functioning as a basis for corrective or improvement-oriented action. In that case planning is likely to have a more "step by step", incremental orientation, and "goals" or expectations get the function of standards for interpreting evaluative information. The discrepancy between actual achievement and expectations creates the dynamics that could eventually lead to more effectiveness. In cybernetics the cycle of assessment, feedback and corrective action is one of the central principles.

Evaluation—feedback—corrective action and learning cycles comprise four phases:

- measurement and assessment of performance;
- evaluative interpretation based on "given" or newly created norms;
- communication or feedback of this information to units that have the capacity to take corrective action;
- actual and sustained use (learning) of this information to improve organizational performance.

In the concept of the learning organization procedural and structural conditions thought to be conducive of this type of cycles are of central importance. Examples are: encouragement of openness and reflectivity, recognition of the importance of exploring different viewpoints and avoiding the defensive attitudes against bureaucratic accountability procedures (Morgan 1986, p. 90).

When the cybernetic principle is seen as the basic regulatory mechanism there is room for autonomy and "self-regulation" at lower levels in the system. Although there may be strong external control over outcomes, "processes" at lower level can

have considerable freedom. This is a particularly helpful phenomenon in education systems, given the usually large degree of professional autonomy of teachers, and tendencies to increase school autonomy as a result of decentralization policies. The functioning conditions of a learning organization are based on the one hand, on the autonomy left to schools to organize methods such as differentiated pedagogies, and on the other hand, on the ability to produce institutional knowledge.

Creating Market Mechanisms: Alignment of Individual and Organizational Rationality

A central assumption in the synoptic planning and bureaucracy interpretation of the rationality paradigm is that organizations act as integrated purposeful units. Individual efforts are expected to be jointly directed at the attainment of organizational goals. In the so-called political image of organizations (Morgan 1986, Chap. 6) this assumption is rejected, emphasizing that "organizational goals may be rational for some people's interests, but not for others" (ibid., p. 195). The fact that educational organizations consist of relatively autonomous professionals, and loosely coupled sub-systems, is seen as a general condition, stimulating political behaviour of the members of the organization.

In public choice theory the lack of effective control from democratically elected bodies over public sector organizations marks these organizations as being particularly prone to inefficient behaviour, essentially caused by the leeway that is given to managers and officers to pursue their own goals besides serving their organization's primary mission.[2]

Public choice theory provides the diagnosis of instances of organizational ineffectiveness, such as goal displacement, overproduction of services, purposefully counterproductive behaviour, "make work" (i.e. officials creating work for each other), hidden agendas and time and energy consuming schisms between sub-units. When discretional leeway of subordinate units goes together with unclear technology, this too adds to the overall nourishing ground for inefficient organizational functioning; see Cohen et al.'s (1972) famous garbage can model of organizational decision-making. Not only government departments but also universities are usually mentioned as examples of types of organizations where these phenomena are likely to occur. Market mechanisms and "choice" are seen as the remedy against these sources of organizational malfunctioning.

Notes of criticism that have been made with respect to the propagation of choice are that parents' choices of schools are based on other than performance criteria, that "choice" might stimulate inequalities in education (Hirsch 1994) and that completely autonomous primary and secondary schools create problems in offering a common educational level for further education (Leune 1994). Apart from

[2]A more extensive treatment of the implications of public choice theory for school effectiveness research is given elsewhere, Scheerens (1992, Chap. 2).

"choice", private schooling, funding formulas, and school autonomy are phenomena that have been addressed in educational effectiveness research associated with market mechanisms (compare the section on cost-effectiveness and the issues listed in Table 1.1).

Self-organization

The concept of *autopoiesis* originates from biology and, in this discipline, was coined to refer to the process of generation and regeneration of living organisms. Attempts were made to "translate" the use of this term to the fields of sociology and law (Luhmann 1995, cited by Kickert 1993). The term was also used in a more metaphorical sense in organizational science to indicate a specific "image" of the functioning of organizations (Morgan 1986). In this context the term was used as the antipode of the common view on organizations as open systems. Instead of "adaptation" in order to survive in case of environmental changes, "closure", autonomy and "self-reference" were seen as key mechanisms of organizational functioning. In more recent interpretations (cf. Kickert 1993) this kind of "closure" is not altogether seen as the complete opposite of an "open" systems view. Instead the organization is seen as taking a "self-centred" view of the external environment; the organization perceives and uses the environment as a projection of itself. Ultimately, in organizational and public administration interpretations, *autopoiesis* is a term that is seen as a helpful metaphor to come to grasps with processes of self-organization of (semi-) autonomous systems.

For organizational study and analysis the theory of self-organization and "autopoiesis" has the following implications:

- recognition of the importance of autonomy, stability and organizational identity; a perspective that bears correspondence with views on school improvement emphasizing cultural aspects in which the "theories in use" of actors have a central place (van den Berg et al. 1999);
- recognition of organizational phenomena aimed at the protection and "buffering" of the existing state of affairs against external influences;
- explicit reservations with respect to external and hierarchical control;
- reservations with respect to the assumption that organizations are goal-, productivity- and efficiency oriented; these criteria are seen as secondary as compared to maintaining the existing social ordering within the organization.

With respect to school improvement and school effectiveness the perspective of *autopoiesis* can be seen as a basis for explaining resistance to change and less "intrinsic" interest for enhancing effectiveness. It could also be seen as a philosophy that underlines the importance of available concepts and cultural preferences of key actors enforcing the status quo in organizations, which defy "easy" transformations. The *autopoiesis* perspective gives reason for less optimism concerning the adaptation of a "culture that favours improvement" and the same can be said with respect to the expected impact of external pressures and the application of rational

techniques like goal setting and assessment. *Autopoiesis* offers a more evolutionary than rational planning perspective. To the extent that its premises are operationally verifiable, it could have an important function in understanding realistic scope and tempo of school improvement.

Making up the Balance on Planning Models and the Degree of Higher Level Structuring

Clearly, of these four planning theories, synoptic planning is strongest in pre-structuring from higher levels. As was noted in previous sections, there might be reason for a revival of national curricula, externally developed lesson plans and adaptive, technology enriched learning environments. But these approaches would not preclude leeway for interpretation and creativity in adaptation to local conditions, nor would they deny seeing learning ultimately as a self-regulated process. In technology enriched learning environments scaffolding can be varied over time, and gradually withdrawn, to provide increased openness for independent problem solving. Another asset of constructivism, the identification of meta-cognitive strategies, could be applied by teaching such strategies in a direct and structured way, next to providing opportunity to apply them in more open assignments. The most contested aspect of the application of synoptic planning approaches in education is the fear of resistance from teachers. This fear may be more warranted in some cultures than in others, but the positive experience with Comprehensive School Reform in the USA cannot be overlooked (Slavin 1996).

Retroactive planning and application of the cybernetic principle in accountability, monitoring and inspection approaches leaves explicit freedom on process dimensions but centralized control on outcomes. As stated before, this approach leaves more autonomy at lower levels, and is therefore less prescriptively controlling, as compared to synoptic planning. Perhaps its focus on outcomes is the most important effect, in stimulating schools to be more result oriented.

Application of market mechanisms shifts the focus of control from vertical to horizontal, by inducing the "direct democracy" of customer control. Since choice, in most practical situations, appears to be quite circumscribed and limited, markets are hardly ever real markets in primary and secondary schooling. What remains as a stimulus for more effective functioning would be largely the same as with the retroactive planning approach: greater result orientation because of more formative and summative assessment. But perhaps community involvement in schooling, parental involvement and discretion of the local community over school functioning deserve to be studied in their own right, not only in association to the idea of creating market mechanisms. The success of Scandinavian school systems, Finland in particular, might be partially explained by the strong involvement of local authorities and the local community in school functioning.

Self-organization draws attention to the importance of entrance states of the units that interact within a system. This underlines the importance of composition effects. Student body composition as well as the composition of the teaching force. When

active policies interact with atmosphere resulting from compositional effects the ecology of schooling is affected. The one kind of meta-control mechanism that could act upon compositional effects and ecology, would be the control of admission, selection as well as the matching (grouping) of students and teachers.

Contingency, Complexity and the Limitations of International Comparison

When describing the four object models the emphasis has been on effective approaches at each level, and in the current analysis of the vertical integration of the four models (Fig. 1.4) the focus has been on vertical across level relationships and general theoretically embedded mechanisms that have an interpretation at each level. Implicitly the focus has been on linear interpretations of "main effects" and not on "interactions" and more complex relationships. Yet, "adaptivity" is to be seen as an important phenomenon at all levels. At the micro levels of teaching and learning adaptive teaching means that the choice and presentation of material is adapted to characteristics of the learner, like his or her ability levels, cognitive style and motivational profile. Instructional research, following the Aptitude Treatment Interaction (ATI) paradigm has identified some relevant interactions; perhaps the most robust one claiming that structured teaching works better with low ability students. At the level of school organization one could agree on the principles of "contingency theory", which states that there is not one effective way to organize. What is effective depends on specific given antecedent conditions, such as the stability of the environment, the nature of the organization's primary process, and, probably most important for schools, on characteristics of teacher and student body composition. This kind of contingency thinking can also be applied at the level of national educational systems, where these are expected to fit important characteristics of the society at large.

In research on educational effectiveness and productivity there has always been a strong interest in establishing robust factors that are "generalizable", across subject matter domains, sub-populations of students, and even national contexts. Of the factors that Walberg (1984), identified, on the basis of massive meta-analyses he said that they showed that "what is good for the goose, is good for the gander". In terms of research strategy both aims: finding generalizable factors on the one hand and showing relevant interactions appear to be perfectly legitimate. A pragmatic approach could consist of taking on as much adaptivity and contingency as considered feasible, and then find out whether associations come out as main effects or interaction effects. More specifically, the conceptual framework is to provide the following handles for taking into consideration adaptivity and contingency:

– the identification of given antecedents at different levels, for example, the socio economic status of students, allows for grouping on these variables and analysing interactions with malleable variables, e.g. checking whether structured

teaching approaches work well for students in the lowest 25 % of the SES distribution;
- the identification of composition variables, such as average SES per school, followed by analysing multi-collinearity and "joint effects" with malleable school variables, such as the level of material resources of a school;
- making deliberate choices concerning variables that are expected to work across disciplines, such as, for example effective learning time, and variables that have a specific subject matter tied interpretation; e.g. teacher's pedagogical content knowledge in mathematics and learning to learn strategies in reading.

Despite such feasible ways to consider adaptivity and contingency to some extent in international comparative assessment studies, these studies are not likely to allow for taking into consideration additional complexity, in the form of higher order interaction effects, non-linear relationships and recursive relationships (e.g. feedback loops), unless data collection would be further expanded. To the degree that we are not able to define all relevant contingency conditions, limitations in the comparability of national educational systems have to be faced.

Choice of "Process" and "Input" Variables, Depending on the Kind of Outcomes That Are Measured

Output indicators might be further differentiated into direct outputs, based on student achievement test scores, attainment outcomes, in the sense of success rates, dropout rates, and impact indicators that would define societal success of people who have attained certain levels of schooling. These three different kinds of output indicators are summarized in Table 1.2.

The choice of input and process indicators for a specific study would depend on the kind of output indicators that are measured. Globally one could say that inputs and process variables should be defined at school and classroom level, when direct outputs in the sense of achievement tests are measured, while macro-level system characteristics seem to be more appropriate in the case of societal impact indicators; attainment indicators could be associated with school level as well as system level input and process factors.

Direct outputs in terms of educational achievement measures could be rank-ordered on a scale that runs from curriculum-tied knowledge and skills, to competencies that border on personality characteristics, see Table 1.3.

To the degree that educational outputs tend to be lower on the above continuum, the relevance of curricular input and process variables becomes debatable. In the PISA study, for example, there has been relatively little attention for describing national curriculum contexts, until quite recently. In more recent waves of the study (2009 and 2012) opportunity to learn specific sub-components of the literacy domains has become incorporated.

Table 1.2 Overview of different kinds of output indicators

Main categories of outcome indicators	Sub-categories	Technical issues
Output indicators	Achievement measures	
	– Subject matter based	Value-added effect measures; growth curves
	– Literacy (reading, mathematical, scientific)	Assessment methodology (ranging from multiple choice tests to authentic assessment)
	– Competencies (e.g. learning to learn)	Criterion versus norm-referenced testing
Outcome/Attainment indicators	Attainment measures	
	– Graduation rates	Controlling for selection oriented school policies
	– Proportion of students graduated without delay	
	– dropout rates	
	– Class repetition rates	
Impact indicators	Social participation rates	
	– (for each attainment level) % of employed at a certain job level	Availability of national educational and labour market statistics
	– % of unemployed	
	– (for lower school levels) % enrolled in follow-up education	
	– Degree of social participation (social capital)	Appropriate measures of social capital and adult literacy
	– Adult literacy rates	
	– Average income, for each attainment level; earning differentials	
	– Skill shortages and surplus	

Table 1.3 A continuum of educational outcomes, running from highly content bound to personality dependent

– Outcomes as measured by tests included in textbooks
– Outcomes as measured by implemented school curricula (teacher developed)
– Outcomes as measured by tests based on the intended national curriculum
– Outcomes as measured by international tests covering the common core of a range of national curricula, e.g. TIMSS
– "Literacy" tests, aimed at measuring basic skills in reading, mathematical and scientific reasoning, e.g. PISA
– Competencies as multi-facetted dispositions of individuals, including cognitive, motivational and possibly other components
– Personality traits, like internally or externally determined locus of control, independence, general intelligence

Conclusion

The functioning of education as a controlled hierarchical system was schematically described in this chapter. The main features of the resulting conceptual framework can be summarized as follows:

- a multi-level structure, with the individual learner, the instructional setting or classroom, the school, and the national education system as the key hierarchical levels;
- the recognition of direct and indirect effects, both within and across hierarchical levels;
- production processes at each level: learning, teaching, school organization and management and macro-level educational governance, characterized in terms of input, process, outcome relationships;
- recognition of core criteria to judge overall functioning and quality of educational systems, namely productivity, effectiveness, efficiency and equity, which could be specified as specific emphases and relationships among input, process and outcome indicators;
- recognition of the impact of contextual and ecological conditions, next to conditions that are directly malleable by key actors in the system;
- variability in structure and independence in the control of a lower level sub-system by a higher level sub-system;
- major planning and management theories (synoptic planning, retroactive planning, market mechanisms and self-organizations) placed on a continuum running from tight control to total autonomy at lower levels, to be seen as underlying mechanisms of the degree of structure and independence in the running of educational systems;
- results of various strands of empirical educational effectiveness research seen as the basis for detecting the most relevant malleable variables at school, classroom and national system's level.

The framework developed in this chapter is to set the stage for the build-up of Part 1 of the book, in which subsequent chapters specify partial models for instructional/teaching effectiveness, school effectiveness and system effectiveness, and this structure is basically maintained in Part 2, when presenting empirical research evidence at classroom, school and system level. The chapter also touches upon theoretical concepts and interpretations "behind" the empirical research results and implications for policy and practice, issues that will be discussed further in Part 3. Finally, the conceptual framework is offering inroads to understanding "ineffectiveness", a theme that emerges gradually throughout the sequence of chapters.

References

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*(4), 409–429.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly, 17*, 1–25.

Creemers, B. P. M., & Kyriakides, L. (2012). *Improving quality in education. Dynamic approaches to school improvement*. London: Routledge.

de Leeuw, A. C. J. (1990). *Organisaties: Management, analyse, ontwerp en verandering: een **systeemvisie***. Assen/Maastricht: Van Gorcum.

Hallinger, P., & Heck, R. H. (2010). Leadership for learning: Does collaborative leadership make a difference? *Educational Management Administration & Leadership, 38*(6), 654–678.

Hirsch, D. (1994). *School: A matter of choice*. Paris: OECD/CERI.

Kickert, W. J. (1993). Autopoiesis and the science of (public) administration: Essence, sense and nonsense. *Organization Studies, 2*, 261–278.

Leune, J. M. G. (1994). Onderwijskwaliteit en de autonomie van scholen. In B. P. M. Creemers (Ed.), *Deregulering en de kwaliteit van het onderwijs* (pp. 27–48). Groningen: RION.

Luhmann, N. (1995). *Social systems*. Stanford, CA.: Stanford University Press.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B. U., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD. Available from: http://www.pisa.oecd.org/dataoecd/15/20/34668095.pdf

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Morgan, G. (1986). *Images of organizations*. Beverly Hills: Sage.

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Chicago: Aldine-Atherton.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.

Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. *Interantional Journal of Educational Research, 13*(7), 691–706.

Slavin, R. E. (1996). *Success for all*. Lisse: Swets & Zeitlinger.

van den Berg, R., Vandenberghe, R., & Sleegers, P. (1999). Management of innovations from a cultural-individual perspective. *School Effectiveness and School Improvement, 10*(3), 321–351.

Walberg, H. J. (1984). Improving the productivity of American schools. *Educational Leadership, 41*, 19–27.

Chapter 2
Modelling Teaching and Learning

Keywords Effective teaching · Learning situations · Pro-active · Interactive and retro-active teaching strategies · Classroom ecology · Structure and independence in teaching · Structured teaching · Constructivist oriented teaching · Creating learning opportunities

Introduction

In this chapter a first conceptualization of teaching and learning will be given. This will be done from the perspective that, obviously, teaching is expected to stimulate learning, but that it would be a strong overstatement to say that teaching determines learning. Instead learning is seen as a process that is, to a considerable extent, controlled by the learner, and in that sense relatively autonomous. Teaching, in this conception, is more like a "booster" to learning than a complete determinant. Such a view fits within a system's perspective on education as a hierarchical structure, in which there is loose coupling between the higher and lower levels, and considerably autonomy in the "production" process at each level.

Student Learning

As a "product" learning can be defined in terms of increased knowledge, skills and acquired norms and values. More interestingly, for our purposes, are characteristics of learning processes, evoked by situations that are expected to stimulate learning.

The following requirements seem to be essential elements experienced by learning subjects in "learning": situations (i.e. situations that evoke learning):

- awareness of the action demands of a situation, feeling challenged to solve a problem, read a text, accomplish a task;
- sufficient motivation and volitional energy to "do something"

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_2

- interpreting the demand aspects of the situation in the light of previous knowledge, habits, available skills;
- executing cognitive operations, varying from simple memorization to so called higher order processes (see for example the taxonomy of cognitive objectives by Bloom);
- possibly applying meta-cognitive strategies which would imply that the previously mentioned operations are to some extent self-controlled and conscious (examples of such meta-cognitive strategies are being aware of objectives, setting targets, decomposing the learning tasks in sub-tasks, using problem-solving strategies, applying self-evaluations);
- delivering a relative (meaning that it would always be possible to learn more) end "product" of the learning experience that could function as auto-feedback to the learner or inform the outside world, particularly a teacher; incidentally auto-feedback would be an important aspect of self-regulated learning.

These characteristics of learning situations can be studied by means of direct or indirect (video recordings) observations of classroom situations and think out loud methods, (Veenman and van Hout-Wolters 2002).

Some students are better in applying the above-mentioned elements of learning processes than others. This would depend on cognitive aptitudes, motivational dispositions that are related to socio-economic background and gender, on cognitive styles and on preference for certain learning strategies. OECD, Artelt et al. (2003) distinguish three kind of learning strategies: memorization strategies, elaboration strategies (which involves exploring how knowledge learned in other contexts relates to new material) and control strategies. The latter kind of strategies are meta-cognitive strategies for self-regulated learning and involve: "checking what one has learned and working out what one still needs to learn, allowing learners to adapt their learning to the task at hand", ibid, p. 13. The latter strategies are highly correlated with effective learning and achievement.

Part of these student background conditions should be seen as given or antecedent conditions to learning at school, others, particularly the ones subsumed under the label of learning strategies, could be taught, by means of explicit lessons (in applying problem-solving strategies and "learning to learn" lessons), or become the object of normal day-to-day "reflective" teaching, where teachers could give cues and hints impinging on students learning strategies. Running ahead to the section on teaching, it is important to refer to approaches like questioning and formative assessment during lessons to facilitate such reflective teaching.

Figure 2.1 gives a schematic presentation of the variables in student learning; following the basic structure of the general conceptual framework in Chap. 1, Fig. 2.1 and Table 2.1 provide a summary of relevant background and partly teachable dispositions.

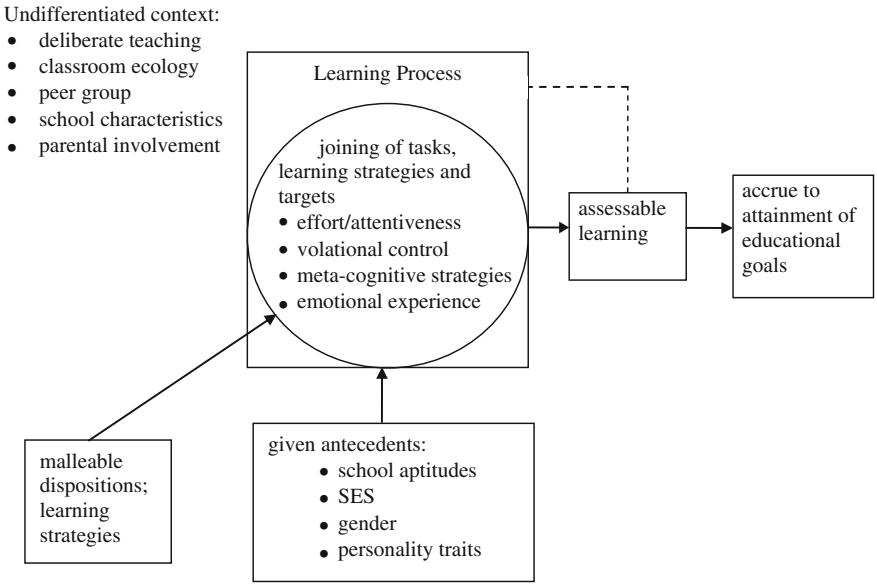

Fig. 2.1 Student learning

Table 2.1 Overview of variables in learning

Background variables	Partly teachable dispositions	Learning processes
General intelligence Scholastic aptitude Socio-economic status Gender Immigrant status Relevant personality traits, e.g. locus of control Cognitive style	Learning strategies: • Memorization strategies • Elaboration strategies • Control strategies • Domain specific strategies Subject matter interest Instrumental motivation Persistence Self-efficacy Subject matter based and general academic self-concept Meta-cognitive knowledge about cooperative reasoning	On-task behaviour Level of engagement Self-report on ongoing learning processes (think out loud procedures)

Teaching and Classroom Ecology

Seen within the framework of a hierarchical presentation of an educational system the first linking issue to be encountered is: what should a teacher do to create effective learning environment and boost learning? In connection to the description of learning situations in the previous section the basic issue could be seen in terms of pre-structuring

the learning situation before and during lessons, and raising the question as to how far pre-structuring should go. This would not be in a setting of total freedom: there is a given organizational arrangement, i.e. a classroom and a certain number of students assigned to that classroom/lesson, hence: ecology (e.g. class size, class composition).

For simplicity's sake, teachers will be seen as the prime "managers" of teaching and learning in classrooms. Later on teaching and learning will partly be attributed to the larger context of the school, characterized by certain organizational properties. The way teachers could influence student learning can be seen in terms of indirect conditions, such as the competencies the teacher brings to the teaching and learning situation, her/his influence on creating the learning environment, curricular choices, text books, media, and assessment tools and direct teaching strategies.

The Elementary Parts of Teaching: Matching of Content Elements, Psychological Operations and Didactic Considerations

In prescriptive formulations, which concern for example the structure of educational objectives, two-dimensional classifications are usually proposed (e.g. De Corte et al. 1973; Bloom et al. 1971). Subject matter content and psychological operations are the two basic dimensions This perspective comes down to breaking down the contents, for e.g. a geography lesson in smaller units, and, for each content unit specifying the cognitive, or affective or psychomotor behaviour/dispositions that should be acquired. On each of these three dimensions a continuum of operations has been specified that ascends in order of complexity. For example: perception of information, recognition of information, reproduction of information, interpretative production of information, convergent production of information, evaluative production of information and divergent production of information. An example of divergent production of information with respect to the geographical climate is the following objective: "On the basis of data on diverse factors that influence the climate of a particular region, deduce certain weather conditions" (De Corte et al. ibid.).

Elements defined by the two basic dimensions (content and psychological operations) could be seen as the demand structures for learning tasks.

In the act of teaching specific presentation forms and media should be applied when introducing the prime didactic elements as described above. The quality of teaching would thus depend on:

- adequate selection of content
- indicating target psychological operations (e.g. cognitive behaviour)
- knowledge about creating tasks (on the basis of the above two elements)
- instructional knowledge; a repertoire of presenting and guiding the execution of learning tasks
- knowledge about students, and typical behaviours for the learning task in question, including frequently made mistakes

The first three characteristics could, in principle, be taken care of outside the direct teaching situation, by curriculum experts and designers of teaching methods. Yet, knowledge about content, about tasks and about student thinking would constitute a basic teaching competency that is described as *pedagogical content knowledge* by Baumert et al. (2005) who found evidence of considerable impact of this variable on student achievement.

According to Baumert et al. ibid, pedagogical content knowledge is a combination of instructional knowledge, content knowledge, knowledge about student thinking and knowledge about tasks as independent tools; it involves in the case of mathematics as the subject matter area as given below:

- Knowledge About the Instructional Potential of Mathematical Tasks
  - local knowledge about tasks, e.g. multiple solutions
  - orchestration of tasks into instructional sequences
  - cognitive demand of mathematical tasks (e.g. multiple steps in modelling, complexity of language)
- Knowledge of Creating Mathematical Meaning in Classroom Interaction
  - multiple representations and explanations
  - cognition of mathematical representations
  - fast recognition of mistakes
  - making use of critical incidents (maintaining the level of cognitive complexity, keeping students responsible for their learning)
- Knowledge About Students' Conceptions and Students' Thinking
  - recognition of misconceptions
  - recognition of typical mistakes
  - knowledge about typical difficulties

In practical terms these aspects of pedagogical content knowledge are manifested not only in the curriculum and in the lesson preparation by teachers, but also in the realm of actual interactive classroom teaching and in monitoring and assessing student activities and work. Interestingly Baumert and his colleagues have worked out the characteristics of pedagogical content knowledge in the form of assessment type of instruments, using computerized tests and video clips to assess teacher competencies in specific content areas, in their case mathematics.

Table 2.2 provides a schematic overview of pro-active, interactive and retro-active phases in the structuring of teaching, based on content and psychological operations as the two basic dimensions.

Combination of content units and psychological elements, bearing in mind pedagogical and didactic considerations, can be seen as lying at the heart of instructional sub-disciplines as:

- curriculum development
- teacher training and teacher professional development
- lesson preparation

Table 2.2 Stages in preparing, executing and evaluating the teaching act

Content dimension	Psychological dimension
– Decomposition of content in sequences that represent the structure of the subject matter area	– Taxonomy of cognitive, affective and psychomotor operations that reflect increasing complexity
COMBINE BOTH DIMENSIONS IN	SEQUENCES OF INSTRUCTIONAL OBJECTIVES
− Creating tasks and task sequences with pedagogical potential	– Taking into consideration cognitive complexity and emotional meaning of tasks
COMBINE BOTH IN	LESSON PLANS AND SCRIPTS
− Actual teaching in which multiple representations and explanations of content elements are given	− Taking into consideration possible misconceptions, typical difficulties and frequently made mistakes
COMBINE BOTH IN	TEACHING
− Constructing content elements for the development of items for formative and summative assessment instruments	− Adding representations of expected psychological operations, with different degree of complexity to each content element of item
COMBINE BOTH IN	ITEM BANKS AND TESTS IN WHICH DIFFICULTY LEVEL AND ABILITY ARE IDENTIFIABLE DIMENSIONS

– actual teaching
– educational assessment
– providing feedback on the basis of assessment

Having identified a common structure in all these sub-disciplines, it leads to the following conjectures:

– consistency between these sub-disciplines for a specific educational objective should be seen as mutually reinforcing; the concept of *opportunity to learn* is a well-known example, as it stresses the consistency between didactic elements that are taught and tested;
– in case of omission of one or several of these sub-disciplines the remaining ones could compensate and keep the basic structure of good quality teaching intact.

Educational systems are likely to vary in the emphasis that is given to specific, singular, instructional sub-disciplines, or to specific combinations of these sub-disciplines. Centralized curriculum had gone out of fashion but is having a resurrection in the USA with the initiative of the common core curriculum. The school improvement movement, initiated in the writing of Matthew Miles and other authors (Miles 1998) emphasized bottom-up development, while a counter-movement can be seen in the work of authors who critically address the school improvement approach as an "ownership paradigm", (Muijs and Reynolds 2001) according to which each school is expected, more or less, to reinvent the wheel.

The use of assessment, particularly "summative assessment" for accountability purposes is quite contested in many countries, but very central in educational

reform in others. Each of the sub-disciplines can be seen as an area of leverage for educational reform and improvement.

Special attention should be given in this context to technology-enriched learning environments. In these applications content and psychological elements, as well as considerations about didactic strategies are integrated in a very explicit way. In the earlier application of computer assisted instruction (CAI) a strong emphasis was given to the presentation of subject matter and the provision of examples in the practice of basic skills in reading, spelling arithmetic and topography. Characteristic of CAI is the use of repetition (drill and practice) and feedback. The application of "branching", allows for taking into consideration individual differences, so that a degree of adaptiveness can be realized. Such programs have been quite successful, particularly for weaker students (cf. van Merriënboer and Kanselaar 2006, p. 287). In intelligent tutoring systems there is more specific attention for psychological and didactic aspects. In the SCHOLAR system (Carbonell 1970, ibid, p. 288) for example, the following four components were integrated: a domain or expert model (core of knowledge as well as frequently made mistakes), a student model, which represents the current level of expertise of the student, an instructional model (e.g. conducting a Socratic dialogue) and an evaluation and feedback model. In later applications dynamic visual representations, and multimedia components are added, while active learning, supported by hints and cues, "scaffolding", are used to support students. Techniques of gradually reducing these kinds of cues, indicated as "fading" can be built into interactive lessons in order to gradually enhance independent learning (De Jong 2006). In this context, Linn et al. (2006), discuss forms of inquiry learning that are supported by visualization techniques and assessment techniques that require students to link and connect ideas in science education.

Structure and Independence in Teaching

Having defined the nature of core instructional elements and the various sub-disciplines in which they are given shape, it is time to turn back to the core idea of seeing teaching as a set of conditions that should facilitate and "boost" student learning. In the section in which the learning process was modelled, student control strategies appeared to be the most effective. In the act of teaching seen as routine control, teaching could be seen as compensating for lack of student control in learning. Teaching as meta-control could address meta-cognitive strategies, by which student control strategies would actually be taught to students.

In a way student control strategies are the pendant of the main features of "structured teaching" and direct instruction, where it is the teacher who actively orders and controls the teaching and learning situation. When putting these two orientation s, with structured teaching on the one hand, and students effectively employing control strategies next to one another on the other hand, the following types of associations can be discerned:

- structured teaching happens as a *substitute* for student control strategies
- structured teaching happens as an *additional support* for student control strategies
- structured teaching happens as a *model and example* to enhance student control strategies (meta-cognition)
- structured teaching happens as a *suppressor* of student control, because students are not given sufficient leeway to develop and manifest this behaviour themselves.

Weaker students in primary and secondary education are more likely to benefit from the first two alternatives, whereas the last two alternative combinations are more probable when dealing with better students in secondary education (where obviously the third alternative is a positive and the fourth is a negative example).

The above interpretations suggest a resolution of the distinction between structured and more open, discovery-oriented teaching approaches, by making it conditional on student aptitudes. Doing so would bring the question concerning structure or independence in the realm of adaptive teaching and research in the tradition of aptitude treatment interaction research (ATI, cf. Cronbach and Snow 1981). Although taking this latter perspective seems to be the more plausible one, it still makes sense to contrast the traditions of structured teaching, mastery learning and direct instruction, with "constructivist ideas", about teaching and learning.

Direct Teaching

Doyle (1985), considers the effectiveness of direct teaching, which he defines as follows:

(1) Teaching goals are clearly formulated.
(2) The course material to be followed is carefully split into learning tasks and placed in a sequence.
(3) The teacher explains clearly what the pupils must learn.
(4) The teacher regularly asks questions to gauge what progress pupils are making and whether they have understood.
(5) Pupils have ample time to practice what has been taught, with much use being made of "prompts" and feedback.
(6) Skills are taught until mastery of them is automatic.
(7) The teacher regularly tests the pupils and calls on the pupils to be accountable for their work.

The question whether this type of highly structured teaching works equally well for acquiring complicated cognitive processes in secondary education can be answered in the affirmative, according to Brophy and Good (1986, p. 367). However, progress through the subject matter can be taken with larger steps, testing need not be so frequent and there should be space left for applying problem-solving strategies flexibly.

Table 2.3 Objectivism versus constructivism, adapted from Jonassen (1991, p. 9)

	Objectivism	Constructivism
Reality	External to the knower	Determined by the knower
Mind	Abstract machine for manipulating symbols	Conceptual system for constructing reality
Thought	Atomistic: decomposable	Gestalt properties
	Manipulates abstract symbols	Imaginative: enables abstract thought
Meaning	Corresponds to entities and categories in the world	Does not rely on correspondence to the world (determined by understanding)
Symbols	Represent reality	Tools for constructing reality

Constructivism[1]

Constructivism is an offshoot of what has been called the cognitive revolution in learning theory. Not only is it opposed to behaviourism but it is also set apart from objectivist instructional applications of cognitivism, such as intelligent tutoring, as these take the expert's knowledge structure as an external entity that is to be mapped on the students. Instead, constructivism emphasizes the active role of the learner in constructing knowledge. The real enemy of constructivism is objectivism.

Some of the opposing claims of both positions as stated by Jonassen (1991, p. 9) are summarized in Table 2.3.

Constructivism claims that reality is more in the mind of the knower, but does not go as far as denying external reality altogether (solipsism), however some radical constructivists come very close to a position of complete denial.

The image of student learning that goes with constructivism underlines the active role of the learner. Students are to be confronted with "contextual" real-world environments, or "rich" artificial environments simulated by means of interactive media.

Learning is described as self-regulated with lots of opportunity for discovery and students' own interpretation of events.

Learning strategies, learning to learn and reflecting on these learning strategies (meta-cognition) are as important as mastering content. Different ways in finding a solution are as important as the actual solution itself. Terms like "active learning" (Cohen 1988), "situated cognition" (Resnick 1987) and "cognitive apprenticeship" (Collins et al. 1989) were used to describe student learning.

Next, students are expected "to construct their own meaningful and conceptually functional representations of the external world" (Duffy and Jonassen 1992, p. 11). The teacher becomes more of a coach, who assists students in "criss-crossing the landscape of contexts", looking at the concept from a different point of view each time the context is revisited (Spiro et al. 1992, p. 8). Cohen (1988) used the term "adventurous teaching" for this approach.

[1]This sub-section is a direct citation from Scheerens (1994).

There is less emphasis on structuring goals, learning tasks and plans in advance; goals are supposed to emerge when situated learning takes place and plans are not so much supposed to be submitted to the learner as constructed in response to situational demands and opportunities.

Learning situations must be such that students are invited to engage in sustained exploration (real-life contents or simulated environments). Some authors writing from this perspective state that "transfer" is the most distinguishing feature (Tobias 1991), whereas others mention argument, discussion and debate to arrive at "socially constructed meaning" (see Cunningham 1991).

The role of assessment and the evaluation of students' progress was hotly debated. Radical constructivists took the position that performance on an actual learning task is the only legitimate way of assessment, since distinct "external" evaluation procedures could not do justice to the specific meaning of a particular learning experience for the student.

Others (e.g. Jonassen 1992) concluded that assessment procedures from a constructivist perspective should merely be different: goal-free, rather than fixed on particular objectives, formative rather than summative, and oriented to assessing learning processes rather than mastery of subject matter. Appraisals of samples of products, portfolios and panels of reviewers that examine authentic tasks were also mentioned as acceptable procedures.

In Table 2.4 some of the major distinguishing features of learning and instruction according to the constructivist position are contrasted with characteristics of more traditional instructional models like direct instruction and mastery learning.

Bipolar comparisons such as the one in Table 2.4 run the risk of oversimplification and polarization whilst also constructing "straw men". It should be emphasized that less extreme constructivist views can be very well reconciled with more "objectivist" approaches (cf. Merrill 1991). Also, more eclectic approaches are feasible, as can be seen when more teacher-controlled and learner-controlled instructional situations are used alternately (cf. Boekaerts and Simons 1993).

Table 2.4 Comparison of traditional and constructivist instructional models

Traditional instruction	Instruction inspired by constructivism
Emphasis on basic skills	Bias towards higher order skills
Subject matter orientation	Emphasis on learning process
Structured approach • Pre-specified objectives • Small steps • Frequent questioning/feedback • Reinforcement through high % of mastery	Discovery learning "Rich" learning environments • Intrinsic motivation • Challenging problems
Abstract-generalizable knowledge	Situation-specific knowledge, learning form cases
Standardized achievement tests	Assessment less circumscribed
	Alternative procedures

Input "given" entrance conditions	**Technology**				**Aims (type of outcomes)**
	Throughput				
	Instruction	*learning*	*classroom organization*	*Tasks*	
low aptitudes	direct	reactive receptive	whole class individual	highly structured	basic skills academic
high aptitudes	open	active productive discovery learning	groups	real life problems	higher order processes "real life" knowledge social skills

Fig. 2.2 Degree of structure in instructional technology conditional upon entrance conditions and goals

An important realization is that the optimal position on continua that represent dimensions of direct teaching versus constructivist teaching is likely to depend on "situational" characteristics like: the kind of instructional objectives that are aimed for and the aptitude of the learners. This last idea is in accord with the research approach referred to earlier as research into aptitude treatment interactions. It also encompasses the idea of adaptive instruction. A general, research supported idea being that weaker students benefit more from more structured teaching approaches.

These ideas are expressed in Fig. 2.2.

Constructivism has helped in enforcing the notion that in education the learner is the core producer of learning, and that teaching is more like an add-on or a boost to learning. More open, discovery kind of learning situations, where students are invited to self-regulate their learning, are usually seen as logical implications for teaching. But here, perhaps the step from description to prescription is too drastic. Strictly speaking one could say that even in a very structured teaching and learning situation, where, for example, worked out solutions to mathematical problems are used as inputs, learning could only take place if the learner could re-construct this stimulus material in a way that would allow him or her to apply to solution to other problems, related problems or practical applications; the ultimate moment of "construction" being the examination assignment or test item where the student would have to demonstrate his or her knowledge. Learning remains a self-directed constructive process even in the case of very structured teaching. Research evidence

in which structured and open teaching approaches were compared is relatively scarce; in the cases where it took place the evidence used to be more in favour of structured approaches (see previously given references and van der Werf 2005), more recent results, to be discussed in subsequent chapters, show both approaches as about equally effective. A further remark about constructivist teaching could be that it might give rise to structured approaches in teaching meta-cognitive strategies and learning from mistakes (see the section on feedback, below). More importantly is probably to get away from pitting these two kinds of instructional strategies against one another. As Fig. 2.2 illustrates, opting for more structured versus open learning situations could be seen as contingent on the learning objectives and the ability level of students. As mentioned before, different emphases can also be projected in the course of one lesson or the learning of a specific complex task. This is illustrated in computer-enriched learning environments, such as the SIMQUEST package in which computer simulations are integrated with supportive cognitive tools (De Jong 2006). The author discusses the option of a gradual withdrawal of cognitive tools, indicated as "fading", where the learner gradually takes over the learning process. Varying different degrees of support and scaffolding in the course of a lesson could show a similar moving from structured to more independent learning forms.

Pro-active and Retro-active Regulation in Teaching: The Case of Performance Feedback

Among the set of educational sub-disciplines that were defined in the previous section, curriculum development, teacher training and teacher professional development and lesson preparation could be said to have a *pro-active* orientation. Of actual teaching one could say that it has an *interactive* orientation while assessment and providing feedback have *a retro-active* orientation.

In this section pro-active and retro-active regulation will be compared; it should be noted that pro-active and retro-active regulations have a clear interpretation in core processes of school management and system-level governance as well.

The ideal of "synoptic" planning is to conceptualize a broad spectrum of long-term goals and possible means to attain these goals. As such it contains the basic logic of social engineering, and planned change, in our case design of teaching and learning situations. In models of planned change the various aspects of synoptic planning are usually structured as phase models; the following description of the different phases is partly based on Ackoff (1981, 74, 75).

In a *first phase* there is a reflection on values and normative aspects that should be attained through specific treatments or specific organizational behaviour. This first phase can also be taken as the phase of defining the problem domain.

In the *second phase* ends planning takes place in the sense that goals and objectives are specified.

In the *third face* means-planning takes place, where, ideally, there should be a rationale for selecting the means (examples in education are results of empirical educational effectiveness or practical experience on "what works" in education).

In a *fourth phase* resource planning is focused at determining "what resources will be required, when they will be required, and how to obtain those that will not otherwise be available" (ibid., 75).

In a *fifth phase* design of implementation and control determines "who is to do what, when, and where, and how the implementation and its consequences are to be controlled, that is, kept on track" (ibid., 75).

In a *sixth phase* (which, by the way, is not specifically mentioned by Ackoff), monitoring and evaluation, which can be seen as part of the control processes, are used for feedback and possible modification of means, goals or even values.

The feedback mentioned in this last phase turns the sequence in steps into a circle that can go on and on. Many authors, including Ackoff, do not take the sequence of phases too seriously and say in fact that they may take place in any order. Others, however, see the way one "steps into" the planning, implementation and feedback circle as non-trivial. Borich and Jemelka (1982) see the planned change process as society's attempts to "maintain equilibrium when the system threatens to become disadvantageously influenced by forces whose effects were previously neglected or would have been difficult to predict" (ibid., 216). They see a qualitative difference, however, in two ways of regaining equilibrium. The first being the traditional one where goals are formulated to determine behaviour, and which one could see as a pro-active orientation (J.S.) the second emphasizing that behaviour provides impetus for goals, which they see as a more *retrospective* orientation. They illustrate the difference in these two orientations with a citation from Weick (1969):

> This sequence in which actions precede goals may well be a more accurate portrait of organizational functioning. The common assertion that goal consensus must occur prior to action obscures the fact that consensus is impossible unless there is something tangible around which it can occur. And this "something tangible" may well turn out to be actions already completed. Thus it is entirely possible that goal statements are retrospective rather than prospective (Weick 1969, The Social Psychology of Organizing, Addison-Wesley, p. 8).

In a retro-active regulation of teaching the assessment instrumentation, for example a large item bank, takes the place of the intended curriculum, and assessments that have content validity with respect to the intended curriculum, are to be seen as tools to adapt teaching on the basis of assessment. In the ideal situation of an exhaustive item bank, "teaching to the test" could be seen as a fully legitimate and recommendable activity. This kind of assessment would usually be seen as formative assessment, implying that assessment would be expected to feed into ongoing teaching activity and lead to adaptation of these practices. When comparing formative and summative assessment, this could be seen as a relative difference, determined by two criteria:

- the relative duration of the feedback cycle, short in formative and long in summative;
- the instrumental or learning-oriented nature of the feedback in formative evaluation, with a stronger emphasis of judgmental aspects in summative evaluation.

The distinction should clearly be seen as relative concerning both criteria. It is difficult to do away with all judgmental "ego oriented" interpretations in formative evaluation, and some authors emphasize the instrumental improvement oriented impact of summative assessment (cf. Bell 1998).

The key mechanism linking assessment and teaching, expected to be reflective and adaptive by using the assessment information, is feedback. The term feedback stems from control theory, with the functioning of the thermostat as the classical example to illustrate it. When the measuring device indicates that the room temperature is below a certain level, the regulating mechanism switches on the heating (De Leeuw 1990, p. 126). In systems theory feedback loops are seen as positive when the loop exhibits self-reinforcing behaviour, (for example when good results increase positive expectations about students' learning, which, in its turn leads to setting higher standards, a more optimistic, achievement oriented climate, more self-confidence and achievement that is further increased) and negative when a loop exhibits goal-seeking behaviour. An example of the latter would be a teacher needing to increase his or her energy in keeping order, when the students' behaviour worsens (Clauset and Gaynor 1982).

In a review of the impact of formative assessment Black and Wiliam (1998) conclude that, across the board, formative assessment and feedback are positively associated with student achievement. However, it is often difficult to separate the impact of assessment-feedback from other regulatory mechanisms that are also active. This is illustrated in their analysis of feedback within the framework of mastery learning, a form of structured teaching comparably to direct teaching, as described in the previous section. Likewise, in reviews and meta-analyses effects of quantitative and qualitative aspects of feedback are sometimes not sufficiently separable.

Key elements of mastery learning are:

- The learner must understand the nature of the task to be learned and the procedure to be followed in learning it.
- The specific instructional objectives relating to the learning task must be formulated.
- It is useful to break a course or a subject into small units of learning and to test at the end of each unit.
- The teacher should provide feedback about each learner's particular errors and difficulties after each test.
- The teacher must find ways to alter the time some students have available to learn.
- It may be profitable to provide alternative learning opportunities.

Table 2.5 Positive and negative conditions affecting effective use of feedback

Positive conditions	Negative conditions
• Feedback not just about standard attainment but also instrumental information, to the extend that feedback recipients can actually use this information	• Feedback exclusively about standard attainment (stimulates a judgmental rather than an instrumental application of feedback)
• The above condition can reinforce the task-related (or *perceived* task related) nature of the feedback	• Feedback is taken "personal" and purely judgmental
• Standards experienced as realistic	• Standards perceived as unrealistic
• Feedback appeals to intrinsic motivation	• Feedback appeals to extrinsic motivation (implication: doubts about incentive schemes in education)
• Negative feedback is superior in stimulating incremental task-related learning progress	
• Positive feedback stimulates overall motivation if received from credible and relevant sources	

Sources Among others: Kluger and DeNisi (1996)

– Student effort is increased when small groups of two or three students meet regularly for as long as an hour to review their test results and to help one another overcome the difficulties identified by the means of the test (McNeill 1969, cited by Black and Wiliam 1998, p. 40).

Although research reviews and meta-analysis of mastery learning usually show positive associations with achievement, the effect sizes are higher when teacher-produced tests as compared to standardized tests are used, when programs are teacher-paced rather than student-paced, and when students are relatively younger. Characteristics of feedback given in the context of mastery learning are that feedback is based on comparing current achievement against a "mastery" level of achievement that such feedback is given rapidly and intended to be diagnostic, and, finally, that students are given the opportunity to discuss with their peers how to remedy any weaknesses.

In search for further insights into the specific characteristics of effective feedback from student assessment, Table 2.5 gives conclusions based on a review by Kluger and DeNisi (1996) on performance feedback (not limited to educational settings).

Feedback

An important characteristic of effective performance feedback is that it is task related or "instrumental", rather than ego related or judgmental.

The idea of instrumental feedback assumes that targets are identified as learning gaps and that there are ideas about mechanisms, means or techniques to bridge learning gaps. Experiencing of learning gaps is closely related to the role of standards

and achievement expectations in teaching. Research on standard setting points out that learning gaps should neither be unattainably high nor be low (cf. De Vos 1989). The assumption of instrumentality and mechanisms to close learning gaps is closely related to matching task characteristics to psychological operations of learners, and knowledge about frequently made mistakes. The relevance of taking task characteristics into consideration, when providing feedback, is highlighted by Black and Wiliam, ibid, p. 52. They say that "feedback appears to be most successful in 'heavily cued' situations, such as are found in computer-based instruction and programmed learning sequences, and relatively less successful in situations requiring 'higher order'—thinking such as unstructured tests and comprehension exercises". They also report evidence that feedback related to progress seems to be more effective than feedback related to absolute performance levels.

The role of the teacher in providing instrumental feedback is important as well. It seems that here we are back to the discussion on structure and independence that was central in the previous section. Teachers have the choice between providing complete solutions, heavily cued hints towards the correct solution, or an adaptive "scaffolding" response, in simpler terms students receiving as much help as they would need to solve the problem on their own.

Moreover, the discussion on effective feedback from student assessment shows the embeddedness in other aspects of teaching and learning situations that are relevant to enhance teaching effectiveness:

- setting of realistic standards
- careful consideration of the didactic aspects of learning tasks
- deliberate choices with respect to pre-structuring on the one hand and independent activity on the other hand
- motivational aspects, e.g. feedback that is experienced as threatening and solely judgmental
- aspects of cooperative learning, in those applications where feedback is discussed in small groups of students
- the quality of assessments

This embeddedness and association with the core elements of teaching (the combination of content selection, choice of target psychological operations and pedagogical and didactic aspects determined by student characteristics) is also illustrated in the way Mislevy et al. (2002) write about the construction of adaptive assessment. They describe a conceptual assessment framework that comprises student models, evidence models, task models and an assembly model.

Basically pro-active curriculum development and construction of didactic tasks on the one hand and achievement test construction on the other hand, can be seen as two sides of the same coin. De Groot and van Naerssen (1966) in this context have coined the terms didactic and evaluative specifications of educational objectives. As it was said before these two processes could mutually support one another, particularly when consistent, but could also replace one another. As an aspect of devolution in educational governance the evaluation and accountability mechanism sometimes seems to have replaced the pro-active curriculum development

mechanism as major form of regulation. When teaching is viewed from the perspective of boosting student learning the retro-active route of assessment and feedback has the advantage of forcing a close link to (intermediate and final) learning outcomes, and reflection on empirical successes and failures in student learning; including attention for meta-cognitive processes.

Providing Learning Opportunities

Early models on effective teaching, such as the well-known Carroll model (Carroll 1963), focus on effective learning time as a key condition. A second major dimension of providing learning opportunities is the content that is covered and that should be representative of the intended curriculum. In fact the deductive line of consistency between intended, implemented and realized (i.e. the way the curriculum is assessed) curriculum is about instruction being targeted and focussed.

Curricular emphases, which might differ between classrooms, schools, and between national educational systems, and classroom management are the two vehicles that should take care of providing learning opportunities. Since these categories have been well researched and documented frequently, no further elaboration will be given in this context.

Classroom Ecology and Climate

The term ecology refers to an approach in biology that is focused at living organisms' habits, modes of life and relations to their environment. Using this term metaphorically for what could be described in a more neutral and sober sense as the classroom *context* is proposed to express the idea that what we are dealing with here is a set of complex interactions between antecedent conditions, such as class size and classroom composition, habits and routines, institutional rules and norms of behaviour and normative orientations, usually gathered by the term classroom climate.

The extensive literature on the impact of class size on student achievement will not be reviewed here. In most general terms a first major outcome is that class size reductions need to be quite substantive before they have any effect; which makes class size reduction a less efficient strategy for improvement. A second important conclusion could be that considerations about class size are contingent on national educational cultures. Relatively larger class sizes in Asian countries do not appear to be considered problematic by teachers and do not preclude high achievement outcomes (cf. Woesmann 2005).

Classroom composition can be expressed in terms of mean scores on student background variables such as socio-economic background and scholastic aptitude, and in terms of the between student variation on these variables. In the latter case we are in the realm of the issue of homogeneous or heterogeneous grouping.

Composition effects, in terms of mean scores on background variables, are mostly studied as school-level variables (cf. Opdenakker and van Damme 2001; Stanat 2003; Baumert et al. 2005b; Willms 2004; Luyten et al. 2005). Substantial effects of composition in terms of mean SES level are shown in these studies, over and above the impact of these variables measured at student level. In more qualitative terms composition effects could reach a critical threshold, after which they bring about a "Gestalt switch" in school and classroom life. Anecdotal evidence about this phenomenon is the experience of a Dutch primary teacher saying that, within a program of inclusive education, she felt that she could manage two children with Down's syndrome in her class, but not three. Stanat (2003), suggests a critical level of around 40 % of immigrant student in a school, after which dramatic changes could occur. Luyten et al. 2005 discuss the strong correlations they found between school climate and school composition, measured as average SES, and interpret the joined effect of these factors as indicative of students from a more advantaged socio-economic background bringing more disciplined habits and more positive perceptions of school values to the school. This latter interpretation is related to normative orientations of the peer group that may be supportive or hostile towards school and classroom life.

The issue of **homogeneous versus heterogeneous classroom composition** has been studied, among others, by Slavin (1996, p. 164). On the basis of his "best evidence syntheses" he makes the following recommendations:

"Leave students in heterogeneous classes most of the time and regroup by ability only in subjects (reading, mathematics) in which reducing heterogeneity is particularly important". He also recommends that grouping plans should be flexible that teachers should vary their level and pace of instruction to correspond to student performance level, and that when within-class ability grouping is implemented, the number of groups should be small.

The matching of teachers and classes can be an issue of explicit deliberation in school policies. Depending on forms of streaming of students' high and low ability classes, decisions could be made, for example, to match the best, and the most experienced teachers to the classes with the lowest average ability (cf. Monk 1989, 1992).

Classroom climate could be defined in holistic terms as the general atmosphere in the classroom. When further analysed the major facets of a favourable, effectiveness enhancing climate are: a supportive style of interaction in teacher student interactions, achievement orientationachievement orientation, clear disciplinary rules, and good student–student interrelationships. Some of these facets are closely connected with more overtly "managed", "institutionalized" and "planned" aspects of teaching. Disciplinary rules could be said to be institutionalized at the school level and maintaining them is one of the important aspects of classroom management. Teacher support is closely connected to the issue of structure and clarity in teaching and feedback, while achievement orientation is connected to the use of explicit or implicit standards, as targets and as assessment norms. More affective aspects of the quality of interactions might be associated with classroom composition. Enacting disciplinary rules is sometimes connected to a larger realm of

normatively good behaviour at school and in the classrooms. In English speaking countries the term "ethos" is sometimes used for this. The school and classroom context, in this way, can be seen as a micro-world in which aspects of "citizenship" are dealt with in a more or less implicit way.

Motivational Aspects: Stimulating Student Engagement

The principle of variation in presentation and didactic methods during lessons is as old as systematic thinking about teaching. More attention for learning strategies of students stimulated by constructivist ideas, and considerations on variation in structured and independent work during lessons (e.g. Boekaerts and Simons 1993) has enforced the interest in what is sometimes generally indicated as "active teaching". "Activating is a syndrome that is centred around the idea of offering students multiple opportunities of active learning (Slavin 1995). In this sense (active teaching, J.S.) comprises teaching aspects such as cooperative learning, situated learning, discovery learning, peer-tutoring, student experiments, hands-on activities, group work, individual work, individual learning, and student discussion" (Seidel et al. 2005, 129).

Teaching forms that strongly appeal to stimulate students to become cognitively active are tasks that require higher order thinking, "cognitive activation" (really understanding what is taught and analysis of Klieme and Rakoczy 2003) deep understanding of content, meaningful contexts, authentic instruction, relevance of contents, appropriate and high level of language, and, variation of different presentation, are subsumed under the general term "**challenge**", by Seidel et al. ibid., p. 131.

Issues of applying standards in target setting and feedback have important implications for student motivation as well. Standards should be ambitious but not to a degree that they are unattainable. Standards are not only determined by planning acts of the teachers, but also enacted in classroom life by group averages (De Vos 1989). Applying standards in a more implicit way is connected to the issue of **teacher expectations** about students' achievement. A positive attitude in which teachers try to get the best out of all students, also the less capable ones, has always been seen as one of the key characteristics of effective schooling and teaching. Flexibility and an "empiricist" attitude would seem to be important characteristics of good teaching as well. The well-known phenomenon of "Pygmalion in the Classroom", which holds a strong message against too early judgment and stereotyping, should be taken in mind as clearly undesirable behaviour. An empiricist attitude could be served by frequent assessment, so that a priori judgments of teachers, are, so to say, continuously put to the test. The interaction of stereotyping and elaborated and rigid stratification in grouping, be it in ability groups, or school categories is one of the potential explanations of the generally lower performance and equity of strongly categorized educational systems (Luyten et al. 2005).

In Fig. 2.3 a schematic overview of the various aspects of teaching, discussed above is given. Figure 2.3 has the same structure as Fig. 1.1 that was introduced in Chap. 1 and, in reference to Fig. 1.3 in Chap. 1, it gives flesh and blood to the teaching level in the hierarchical framework. It should be noted that, while teacher characteristics have not been discussed in this chapter, they will be treated in Chap. 3. Similarly more detailed descriptions of the variables listed in Fig. 2.3 and Table 2.6 will also be given in Chap. 3.

Table 2.6 lists the key variables, once more, categorized as teacher characteristics, classroom ecology and climate and teaching processes. Teacher characteristics are seen as "given", although most of them should be seen as malleable through training and professional development. Classroom ecology and climate are partly "given" and partly malleable by teachers. Teaching processes, finally, are seen as malleable conditions.

In summary then, we could turn back to the key mechanisms by which teaching is seen to stimulate and facilitate learning. How can teaching stimulate learning?

We have seen that the core of teaching can be broken down in pro-active, interactive and retro-active component of confronting students with cognitively and motivationally challenging learning tasks, which take into consideration

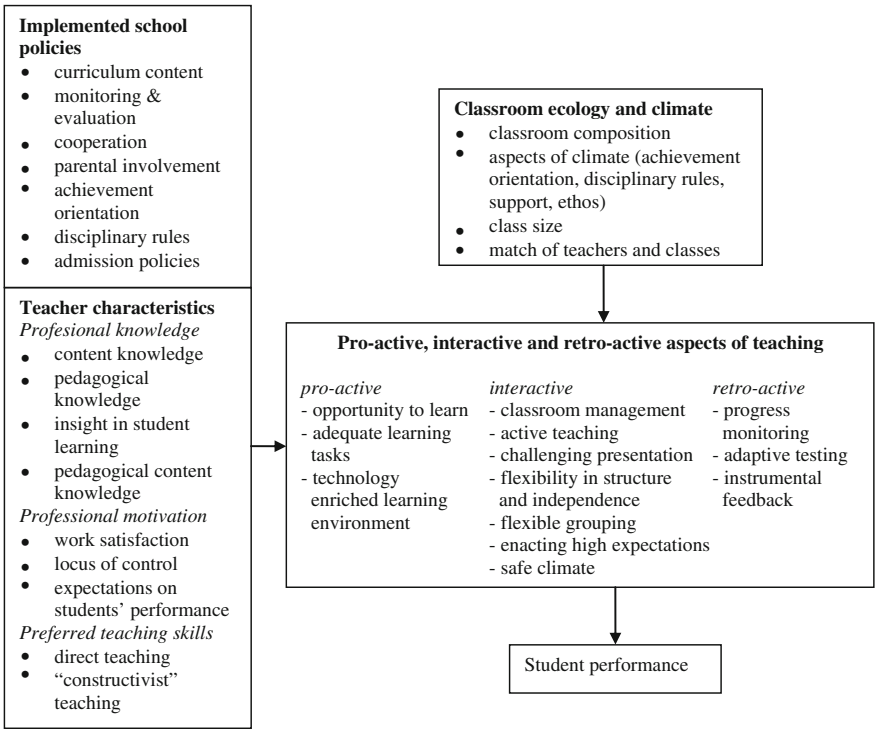

Fig. 2.3 Teaching

Table 2.6 Overview of teaching variables

Teacher background characteristics	Classroom ecology and climate	Teaching processes
Professional knowledge – Content knowledge – Pedagogical knowledge – Insight in student learning – Professional content knowledge *Professional motivation* – Work satisfaction – Locus of control *Preferred teaching styles* – Direct teaching – "Constructivist" teaching	– Class size – Classroom composition (average and heterogeneity) – Match of teachers and classes – Aspects of classroom climate, achievement orientation, discipline, support, ethos – Teacher expectations on students' achievement	*Pro-active strategies* – Opportunity to learn – Selection and design of adequate learning tasks – Technology-enriched learning environments *Interactive strategies* – Classroom management aimed at optimizing active learning time and opportunity to learn – Optimizing structure and independence in teaching – Learning to use learning strategies – Allowing for manageable adaptivity in teaching – Active teaching, diversity in preparation formats – A challenging presentation; cognitive activation; – Enacting high expectations *Retro-active strategies* – Setting realistic motivating standards – Progress monitoring and assessment – Adaptive testing – Instrumental feedback

psychological dimensions of the learner, and try to use knowledge of student learning to support learning processes. Alternating moments of structure and independence determine the specific way this support is given. Core inputs are professional teaching skills, while ecological factors, like class composition, have an important mediating role, as have aspects of classroom management that should culminate in optimized time on task.

Conclusion

This chapter started with a conceptualization of learning as a partly autonomous process determined by the learner. Teaching was seen as stimulating specific aspects of learning processes and creating a general context, in which learning would be enhanced. From an analytic perspective teaching consists of presenting content elements and stimulating certain psychological operations with learners.

In creating teaching and learning situations pro-active, interactive and retro-active aspects were distinguished. Pro-active aspects are related to curriculum and lesson planning, while teaching competencies and the development of those competencies among teachers could also be gathered under this pro-active dimension. The interactive dimension comprises the actual process of creating a learning environment and carrying out teaching acts, while the retro-active dimension is focused at reflection, evaluation, assessment and feedback.

Dimensions of teaching that were seen as central in variations that have been propagated in various teaching models were:

- structure versus independence in teaching and learning situations
- emphasizing pro-active and retro-active aspects
- the distinction between direct teaching activities and creating a classroom ecology
- creating learning opportunities
- stimulating student engagement

In the next chapter a further focus will be given on key factors on which empirical research has concentrated.

References

Ackoff, R. L. (1981). *Creating the corporate future*. New York: Wiley.

Baumert, J., Blum, W., Brunner, M., Jordan, A., Klusmann, U., & Krauss, S. et al. (2005a). *Professional knowledge of teachers, cognitively activating instruction, and the development of mathematical competence (COACTIV)*. Presentation at the meeting of the OECD INES taskforce on teaching effectiveness, Reykjavik, 2005.

Baumert, J., Carstensen, C. H., & Siegle, T. (2005b). Wirtschaftliche, soziale und kulturelle Lebensverhältnisse un regionale Disparitäten des Kompetenzerwerbs. In PISA-Konsortium Deustschland (Hrsg.) *PISA 2003. Der zweite Vergleich der Länder Deutschlands – Was wissen und können Jugendliche?* Münster/New York/Berlin/ München, 2005.

Bell, P. (1998). *Designing for students' science learning using argumentation and classroom debate*. CA, Berkeley: University of California.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evalaution of student learning*. New-York: MacGraw-Hill.

Boekaerts, M., & Simons, P. R. J. (1993). *Leren en Instructie. Psychologie van de Leerling en het Leerproces*. Assen: Dekker & Van de Vegt.

Borich, G. D., & Jemelka, R. P. (1982). *Programs and systems, an evaluation perspective*. New York: Academic Press.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 328–375). New York: MacMillan.

Carbonell, J. R. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems, 11*(4), 190–202.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 722–733.

Clauset, K. H., & Gaynor, A. K. (1982). A systems perspective on effective schools. *Educational Leadership, 40*(3), 54–59.

Cohen, D. K. (1988). Teaching practice … Plus ça change …. In P. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice*. Berkeley, CA: McCutchan.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction* (pp. 453–495). Hillsdale: Lawrence Erlbaum.

Cronbach, L. J., & Snow, R. E. (1981). *Aptitudes and instructional methods: a handbook for research on interactions*. New York: Irvington Publishers.

Cunningham, D. J. (1991). In defense of extremism. *Educational Technology, 31*(9), 26–27.

De Corte, E., et al. (1973). *Beknopte didaxologie*. Groningen: H.D. Tjeenk Willink.

de Vos, H. (1989). A rational-choice explanation of composition effects in educational research. *Rationality and Society, 1*(2), 220–239.

Doyle, W. (1985). Effective secondary classroom practices. In M. J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook*. Washington DC: US Government Printing Office.

Duffy, T. M., & Jonassen, D. H. (1992). *Constructivism and the technology of instruction: A conversation*. Hillsdale, NJ: Lawrence Erlbaum Ass.

Groot, A. D. de, & Naerssen, H. van (1966). *Vijven en zessen. Cijfers en beslissingen; het selectieproces in ons onderwijs*. Groningen: Wolters.

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research and Development, 39*(3), 5–14.

Jonassen, D. H. (1992). Evaluating constructivist learning. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 138–148). Hillsdale, NJ: Lawrence Erlbaum Ass.

Jong, A., & de, C. J. (2006). Computer simulations: Technological advances in inquiry learning. *Science, 312*(5773), 532–533.

Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, et al. (Eds.), *PISA 2000 - Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 333–360). Opladen: Leske+Budrich

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin, 119*(2), 254–284.

Leeuw, A. C. J. de (1990). Organisaties: management, analyse, ontwerp en verandering: een **systeemvisie**. Assen/Maastricht: Van Gorcum.

Linn, M. C., et al. (2006). Inquiry learning: Teaching and assessing knowledge integration in science. *Science, 313*(5970), 1049–1050.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B. U., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD Publishing.

McNeill, J. D. (1969). Forces influencing curriculum. *Review of Educational Research, 39*(3), 293–318.

Merrill, M. D. (1991). Constructivism and instruction design. *Educational Technology, 31*, 45–53.

Miles, M. (1998). Finding keys to school change: A 40-Year Odyssey. In A. Hargreaves, A. Liebermann, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change*. Dordrecht/Boston/London: Kluwer Publishing Company.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19*(4), 477–496.

Merriënboer, J.J.G. van, & Kanselaar, G. (2006). Waar staan we na 25 jaar onderwijstechnologie in Vlaanderen, Nederland en de rest van de wereld? *Pedagogische Studiën, 83*(4), 278–300.

Monk, D. H. (1989). The education production function: It's evolving role in policy analysis. *Educational Evaluation and Policy Analysis, 11*(1), 31–45.

Monk, D. H. (1992). *Microeconomics of school productions*. Paper for the economics of education section of the international encyclopedia of education.

Muijs, D., & Reynolds, D. (2001). *Effective teaching*. London: Paul Chapman.

OECD, Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life. Student approaches to learning. Results from PISA 2000.* Paris: OECD Publishing.

Opdenakker, M., & van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*, 407–432.

Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academic Press.

Scheerens, J. (1994). The school-level context of instructional effectiveness: A comparison between school effectiveness and restructuring models. *Tijdschrift voor Onderwijsresearch, 19*(1), 26–38.

Seidel, T., Rimmele, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction, 15*(6), 539–556.

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston: Allyn & Bacon.

Slavin, R. E. (1996). *Success for all*. Lisse: Swets & Zeitlinger.

Spiro, R. J., Feltowich, P. J., Jacobson, M. J., & Caulson, R. L. (1992). Cognitive flexibility, constructivism and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation*. Lawrence Erlbaum Ass: Hillsdale, NJ.

Stanat, P. (2003). *Predicting achievement of immigrant students: The role of student background at the level of individuals and schools*. Presentation at the 10th biannual EARLI conference in Padova, Italy, August 26–30, 2003.

Tobias, S. (1991). An eclectic examination of some issues in the constructivist-ISP controversy. *Educational Technology, 31*(9), 41–43.

Veenman, M. V. J., & Hout-Wolters, B. H. A. M. van (2002). Het meten van metacognitieve vaardigheden. In F. Daems, R. Rymenans, & G. Rogiest (Eds.), *Onderwijsonderzoek in Nederland en Vlaanderen. Proceedings 29e ORD 2002* (pp. 102–103). Wilrijk: Universiteit van Antwerpen.

Weick, K. E. (1969). *The social psychology of organizing*. Reading, MA: Addison-Wesley Publishing.

Werf, M. P. C. van der (2005). *Leren in het studiehuis: consumeren, construeren of engageren?* Groningen: GION. Rede Universiteit Groningen.

Willms, J. D. (2004). *Contextual effects on student outcomes*. Canada: University of New Brunswick, Canadian Research Institute for Social Policy.

Woesmann, L. (2005). Educational production in Europe. *Economic Policy, 43*(20), 445–504.

Chapter 3
Defining the Key Factors in Instructional Effectiveness

Keywords Teaching effectiveness · Instructional effectiveness · Effective teaching and effective teachers · Classroom management · Teachers' knowledge (content knowledge · Pedagogical knowledge and pedagogical content knowledge) · Opportunity to learn · Effective teaching time

Introduction

In this chapter, the factors that have been addressed in empirical research on teaching and instructional effectiveness will be further defined. The chapter has a substantial annex, in which the core variables of 40 empirical studies are described in more detail.

Once again the basic model of educational effectiveness is used as the ordering framework; cf. Fig. 3.1, which is reprinted once more below to facilitate the use this chapter as a standalone text. Whereas in the previous chapter the key-variables in teaching and instructional effectiveness were generated on the basis of conceptual and theoretical ideas about learning and effective teaching, this chapter remains closer to the variables used in empirical research, as evident from research reviews. Although these two approaches largely coincide, there is not a complete match in the selection of key factors.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_3

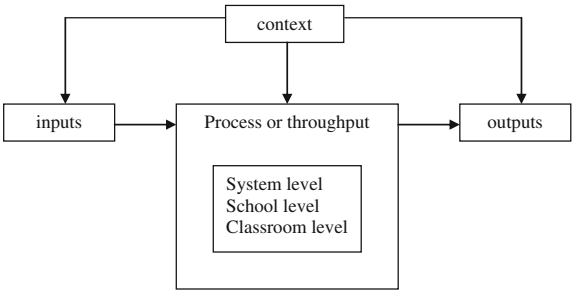

Fig. 3.1 A basic systems model on the functioning of education (reprint of Fig. 1.1)

Review Studies

Numerous reviews studies on school effectiveness and instructional effectiveness,[1] have been published since the late seventies. Early reviews are those by Anderson (1991), Cohen (1988), Dougherty (1981), Edmonds (1979), Good and Brophy (1986), Kyle (1985), Murnane (1981), Neufeld et al. (1983), Purkey and Smith (1983), Ralph and Fenessey (1983), Rutter (1983) and Sweeney (1982). During the nineties reviews were published by Cotton (1995), Creemers (1994), Levine and Lezotte (1990), Reynolds et al. (1993), Sammons et al. (1995), and Scheerens (1992). Scheerens and Bosker (1997) presented the results of a meta-analysis of educational effectiveness research and Scheerens (2000) updated this review by including educational effectiveness studies in developing countries. Research syntheses on educational production functions, as those by Hanushek1997) and Hedges et al. (1994), Scheerens and Bosker's meta-analysis and meta-analyses on instructional factors by Fraser et al. (1987) show negligible impacts of resource related factors and "surface" measures of teacher qualifications (such as highest formal qualification and teacher experience), small to negligible effects of school organizational variables such as educational leadership, coordination, achievement-oriented policy and climate, evaluation practices, time on task and opportunity to learn and medium size effects for aspects of structured teaching, such as providing feedback and reinforcement (Scheerens and Bosker 1997, p. 305). Studies carried out in developing countries show that resource input variables have considerably more impact (Hanushek 1995; Scheerens 2000).

More recently, the "meta–meta analyses" by Hattie, should be mentioned (Hattie 2009; Hattie and Alderman 2012), as well as a new set of review studies by Reynolds et al. (2014), on educational effectiveness, Muys et al. (2014) on teaching effectiveness and Hopkins et al. (2014) on system-level effectiveness and school improvement. The development of a "dynamic model of educational effectiveness",

[1]Although, in this chapter the emphasis is on teaching and instructional effectiveness it would be artificial to exclude review studies that have a certain emphasis on school level conditions, but nevertheless include teaching variables and teaching strategies, defined at school and/or classroom level.

by Creemers and Kyriakides (2008, 2012) provides elaborate multi-level integration of system, school-and classroom-level effectiveness. Finally, important new work with respect to the effects of teacher education and teacher knowledge completes the picture (Baumert et al. 2010; Blömeke et al. 2014).

In this chapter, the focus will be on teaching effectiveness, and to a lesser degree on teacher effectiveness. In subsequent chapters school and system-level effectiveness will be discussed. The review literature will be used to indicate the major developments and emphases over time, in order to categorize and re-define the most relevant variables.

Developments in the Research on Teacher and Teaching Effectiveness

Early Emphases on Personality Characteristics of Teachers

In the sixties and seventies, the effectiveness of certain personality traits of teachers was particularly studied. Medley and Mitzel (1963), Rosenshine and Furst (1973) and Gage (1965) are among those who reviewed the research findings. From these it emerged that there was hardly any consistency found between personality traits of the teacher like being warm hearted or inflexible on the one hand, and pupil achievement on the other. When studying teaching styles (Davies 1972), the behavioural repertoire of teachers was generally looked at more than the deeply rooted aspects of their personality. Within the framework of "research on teaching" there followed a period in which much attention was given to observing teacher behaviour during lessons. The results of these observations, however, in as far as they were related to pupil achievement, rarely revealed a link between teacher behavior and pupil performance (see Lortie 1973, for instance). In a following phase more explicit attention was given to the relation between observed teacher behaviour and pupil achievement. This research is identified in the literature as "process-product studies". Weeda (1986, p. 68), summarises variables which emerged "strongly" in the various studies:

1. *Clarity*: clear presentation adapted to suit the cognitive level of pupils.
2. *Flexibility*: varying teaching behaviour and teaching aids, organizing different activities, etc.
3. *Enthusiasm*: expressed in verbal and non-verbal behaviour of the teacher.
4. *Task-related and/or businesslike behaviour*: directing the pupils to complete tasks, duties, exercises, etc. in a businesslike manner.
5. *Criticism*: much negative criticism has a negative effect on pupil achievement.
6. *Indirect activity*: taking up ideas, accepting pupils' feelings and stimulating self-activity.

7. *Providing the pupils with an opportunity to learn criterion material*—that is to say, a clear correspondence between what is taught in class and what is tested in examinations and assessments.
8. Making use of *stimulating* comments: directing the thinking of pupils to the question, summarizing a discussion, indicating the beginning or end of a lesson, emphasizing certain features of the course material.
9. *Varying the level* of both cognitive questions and cognitive interaction.

Weeda (1986, p. 69) observed that in the study from which these nine teaching characteristics were drawn, there was much criticism regarding methodology/technique. He divided the later research studies focused at instructional effectiveness into two areas:

- pedagogic studies aimed at tracing certain environmental factors and teaching behaviour that can influence levels of performance of certain groups of pupils;
- instructional psychology research aimed at establishing the interaction between teaching variables and pupil characteristics; the so-called *aptitude-treatment-interaction* studies.

Models on Structured Teaching Inspired by the Caroll Model

A next phase in research consisted of process-output studies in which environmental factors and teaching behaviour were studied as influencing levels of performance of certain groups of students and of instructional psychology research aimed at establishing interaction effects between teaching variables and pupil characteristics. The latter studies were known as *aptitude-treatment-interaction* studies.

A central factor within the first area was effective teaching time. The theoretical starting points can be traced back to Carroll's teaching-learning model (Carroll 1963). Chief aspects of this model are:

- actual net learning time which is seen as a result of:
- perseverance and opportunity to learn;
- necessary net learning time as a result of:
- pupil aptitude, quality of education and pupil ability to understand instruction.

The mastery learning model formulated by Bloom (1976) was inspired by Carroll's model.

The findings of the aptitude-treatment-interaction studies were generally judged to be disappointing. There were scarcely any interactions discovered. De Klerk (1985) regarded the fact that the ATI research had failed to reveal any simple interaction between pupil characteristics and instruction method as a challenge to do more refined empirical research on more complex interaction patterns.

Stallings (1985) summarised research literature on effective instruction—in as far as it was concerned with primary education—under the headings: effective net learning time, class organization and management, instruction, assessment and teacher expectations.

When studying net learning time it emerged that simply making the school day longer did not necessarily lead to better levels of performance. More important, ultimately, is how effectively time is spent. Stallings and Mohlman (1981) established that effective teachers spent 15 % of the school day on organisation and management; 50 % on interactive teaching and 35 % on monitoring pupils' work. Aids for an effective use of instruction time include all types of lesson planning. Under the classification *class organization and management* Stallings discusses streaming and maintaining order. Studies on streaming or working with ability groups as compared to whole class instruction indicate that this type of teaching works more positively with the more gifted pupils and that with less able groups—taking the average result of the large numbers of surveys—hardly any effect was found (also according to Kulik and Kulik 1982; Van Laarhoven and De Vries 1987, Reezigt 1993; Slavin 1987). Moreover, from various types of studies it emerged that in classes where there is disruptive behaviour, pupil performance is lower: disruption, naturally enough, is at the cost of effective learning time.

The question what makes good teaching should be looked at on different levels. For direct question-and-answer type knowledge other teaching strategies are called for than for problem-solving and acquiring insight. For learning tasks which greatly depend on memory, a highly ordered and consistent approach is the most effective. Also for the acquiring of insight a clear presentation of the information offered is important, as are questions to check whether pupils have actually absorbed a specific insight. With regard to problem solving, some empirical support is available which shows that it is desirable that pupils take much initiative themselves. Collins and Stevens (1982) mention five teaching strategies to support learning in the sense of problem-solving: (a) a systematic variation of examples; (b) counter examples; (c) entrapment strategies; (d) hypothesis identification strategies; (e) hypothesis evaluation strategies.

From studies on *teacher assessments and expectations* of pupils it seems that self-fulfilling prophecies can occur. If a teacher has once formed negative expectations of certain pupils (s)he is likely to give them less attention and expose them less to more difficult and challenging tasks. Obviously, this is even more of a disadvantage if the initial assessment was a wrong one. Thus, it is imperative that teachers should try and avoid negative stereotyping of pupils (Van der Hoeven-Van Doornum 1990).

In a review of research literature on effective teaching at *secondary school level* Doyle (1985) dealt broadly with the same categories as Stallings, namely "*time on task*" and "*quality of instruction*". Because in secondary education the total teaching spectrum from which a choice must be made is far greater than in primary education, the variable "opportunity to learn" is associated here with the concept of effective net learning time. "Opportunity to learn" is generally understood in the sense of offering pupils a range of subjects and tasks that cover educational goals.

In educational research, opportunity to learn concentrates on the extent to which classroom exercises correspond with the content of the tests for monitoring performance.

As far as the quality of instruction is concerned, there is a stronger emphasis in secondary education on learning higher cognitive processes like insight, flexibly adopting knowledge and problemsolving. Doyle considered the effectiveness of direct teaching, which he defined as follows:

1. Teaching goals are clearly formulated.
2. The course material to be followed is carefully split into learning tasks and placed in sequence.
3. The teacher explains clearly what the pupils must learn.
4. The teacher regularly asks questions to gauge what progress pupils are making and whether they have understood.
5. Pupils have ample time to practice what has been taught, with much use being made of "prompts" and feedback.
6. Skills are taught until mastery of them is automatic.
7. The teacher regularly tests the pupils and calls on the pupils to be accountable for their work.

The question whether this type of highly structured teaching works equally well for acquiring complicated cognitive processes in secondary education can be answered in the affirmative, according to Brophy and Good (1986, p. 367). Doyle also emphasised the importance of varying the learning tasks and creating intellectually challenging learning situations. For the latter an evaluative climate in the classroom, whereby daring to take risks even with a complicated task is encouraged, is a good means. In addition, Doyle discussed the effect of certain ways of working and grouping, including individual teaching and working together in small groups. Bangert, Kulik and Kulik's meta-analysis (1983) revealed that individualized teaching in secondary education hardly led to higher achievement and had no influence whatsoever on factors like the self-esteem and attitudes of pupils.

The results in this summary of reviews and meta-analyses indicate that resource-input factors on average have a negligible effect; school factors have a small effect, while instructional conditions have a small to average effect. There is an interesting difference between the relatively small effect size for the school-level variables reported in the meta-analyses and the degree of certainty and consensus on the relevance of these factors in the more qualitative research reviews.

A Blending of Structured Teaching and Constructivist Principles

In this section the results of some more recent contributions and reviews are briefly summarized, on the basis of work by Anderson (1991, 2004), Brophy (2001),

Baumert et al. (2001), NCS (2002), Muijs and Reynolds (2001), Klieme and Rakoczy (2003), Brophy (2001) and Klieme (2012). In these more recent reviews, a strong corroboration of the main characteristics of effective instruction as laid out in earlier reviews (Scheerens and Bosker 1997) can be discerned. In addition to this consolidation in the knowledge base there are a few additional newer trends. These are the following:

- a reconsideration of characteristics of effective teachers,
- more attention to the teaching of higher order skills, self-regulated learning and "constructivist" approaches,
- a strong re-statement of the fact that teaching is about facilitating learning, by considering learning activities and student engagement.

In the United States, the issue of effective teacher characteristics received much attention, around the turn of the century, in the debate about standards for teaching competency (Darling-Hammond 1999). Empirical studies indicated that subject matter mastery and verbal skills are important assets of teacher effectiveness. In the United Kingdom, McBerr (2000, cited by Anderson 2004) identified 12 characteristics, in the sense of relatively stable traits, associated with effective teachers. Among others he mentions characteristics like flexibility, trustworthiness, and commitment. An interesting feature in Hay McBerr's list is the "drive for improvement". This trait is similar to the "relentlessness" that is emphasized in Slavin's "Success for All" program (1996) and what Anderson (1991) called "zero tolerance to failure". The dimension of confidence is associated with the "high expectations" factor in the school and classroom climate, as one of the frequently identified factors of effective schooling. Klieme and Rakoczy (2003) distinguish a similar dimension, which they call "achievement press" (*Leisteungsdruck*).

There is one other dimension in which more recent contributions return to an aspect that was also present in the very early publications on teaching effectiveness, as Gagne's conception about "the conditions of learning" (Gagne 1971) and the Caroll model, (Caroll 1963). This is attention for student engagement and learning strategies as the ultimate "mediator" between teaching activities and student outcomes. In the OECD study on "student approaches to learning" a range of variables related to engagement is discerned, variables like "self-efficacy", "instrumental motivation" and subject matter interest (OECD 2003). As learning strategies a distinction is made between memorization strategies, elaboration strategies and control strategies. Control strategies have a significant association with reading performance. Control strategies refer to students' ensuring that their learning goals are reached. "These strategies involve checking what one has learned and working out what one still has to learn, allowing learners to adapt their learning to the task at hand", (OECD 2003, p. 13). In a way, these control strategies are the pendant of the main features of "structured teaching" and direct instruction, where it is the teacher who actively orders and controls the teaching and learning situation. Baumert et al. (2001) interpret instruction as an opportunity structure for insightful learning. "This means that instructional materials, task selection and instructional processes are analyzed from the perspective of whether they foster or obstruct active individual

knowledge acquisition. Dimensions of this opportunity structure include the safeguarding of the social action framework by means of appropriate classroom management; pacing and range of learning opportunities (quantity of instruction); general instructional quality, in particular the didactical quality of the structure and realization of the instruction; and the quality of teacher-student and student-student relations."

They go on to say that:

"With respect to general properties of classroom management and the quantity and quality of instruction, robust findings from previous research give a good indication of which aspects of mathematics instruction need to be assessed." (in the context of the OECD PISA-study)

- "Important properties of *classroom management* include: clearly defined rules and procedures, prevention of disturbances, effective responses to critical events and routinization of basic social acts in the classroom.
- Important aspects of *learning opportunities* and pacing (and hence the quantity of instruction) include: learning opportunities with respect to the test items, appropriate ratio of material covered to lesson time (pacing), faithfulness to objectives and relevance of the instructional materials, pressure to perform and interaction tempo.
- The basic properties of *instructional quality* include: level of difficulty, clarity and structure in the presentation of material, adaptivity and individualization of instruction, remediality, participation in instructional activities, monitoring of student activities and general constructivist properties of insightful learning."

The following dimensions of the quality of teacher–student relations are considered: the teacher's ability to motivate students, social orientation and diagnostic competence in the social domain, as well as the students' general satisfaction with their subject teacher.

About the quality of student–student relations in learner groups they propose the following dimensions: cohesion and formation of cliques, competitiveness, mutual assistance, aggression and violation of norms. They also underline the significance of subjective norms with respect to the academic or non-academic orientation of the student body.

Again, in this contribution there seems to be some degree of combination of the "traditional" aspects of structured teaching and constructivist ideas on learning and instruction (Baumert et al. ibid).

Domain Specific Differentiation

In the domain of teaching strategies a difference can be made between general strategies and subject matter, or domain specific strategies. This can be illustrated in particular with respect learning and meta-cognitive strategies. The learning strategies that were discussed in the previous section (memorization strategies,

elaboration strategies and control strategies) are general strategies. In teaching, students can be taught to employ these general strategies and this can be indicated as "learning to use learning strategies".

But it is also possible to consider learning to use learning strategies in a more domain specific interpretation. Seidel and Steen (2005) and Seidel and Shavelson (2007) distinguish domain specific learning strategies in mathematics (math problem solving), science (scientific inquiry) and language (reading and writing strategies).

The recognition of domain specific learning strategies and teaching these to students is theoretically related to the concept of "pedagogical content knowledge", which will be discussed in a subsequent section of this chapter as a specific from of teacher knowledge. Broadly speaking pedagogical content knowledge considers subject specific didactic strategies related to subject matter content. Teaching domain specific learning strategies can be seen as enacted pedagogical content knowledge.

Summary of More Recent Reviews of the Variables in Instructional Effectiveness

The results of these more recent reviews, in the sense of the most important instructional conditions that were referred to, are summarised in Table 3.1. The table includes the main observation categories of a classroom observation schedule, used by the educational Inspectorate in the Netherlands.

The currently most recent overview of key factors in teaching effectiveness is the one by Muys et al. (2014) see Table 3.2.

Contributions by Anderson (2004), Good Wiley and Florence (2009), Klieme (2012) and Muijs et al. (2014), basically corroborated these effectiveness enhancing conditions, with a slightly stronger emphasis on fostering student engagement and stimulating meta-cognition. An attempt at synthesis, based particularly on Good, Wiley and Florence, Klieme and Anderson is provided in Table 3.3.

It can be concluded that research on effective teaching has addressed a relatively consistent set of core concepts. Constructivist ideas have been blended with the more behaviorist-oriented structured teaching models of an earlier phase of development. This is confirmed in the table based on Muys et al. (2014) and in the further structured scheme, developed by Good et al. (2009) and Klieme (2012). The variables that are shown in the various overviews can be seen as specifications of the more basic categories generated in Fig. 2.3. In the next section, *teacher* effectiveness will be considered, before presenting an overview of more operational constructs and variables.

Table 3.1 Summary of reviews and the observation categories of the Dutch Inspectorate

Teaching (Anderson)	Brophy
Enacted curriculum	Opportunity to learn
Classroom physical environment	Curricular alignment
Classroom climate	Supportive classroom climate
Classroom organisation and management	Achievement expectations
Actual teaching	Cooperative learning
Pre-conditions (lesson planning)	Goal-oriented assessment
Communication with students	Coherent content; clear explanations
Stimulating involvement	Thoughtful discourse establishing learning orientations
Dutch inspectorate	Baumert et al. (2001)
• Learning time	Quantity and quality of instruction
• Support in climate	Teacher student relations
• Challenge in climate	Student-student relations
• Structure in teaching	
• Activating students	
• Teaching learning strategies	
• Attainment/teacher focus on attention	
• Classroom organization	

Table 3.2 Summary of effectiveness enhancing teaching variables by Muys et al. (2014)

Teacher effectiveness (Muys et al. 2014)
Opportunity to learn
Time
Classroom management
Structuring and scaffolding, including feedback
Productive classroom climate
Clarity of presentation
Enhancing self-regulated learning
Teaching meta-cognitive strategies
Teaching modelling
More sophisticated diagnosis
Importance of prior knowledge

Adapted from Scheerens (2014)

A Closer Look at Teacher Effectiveness

While *teaching* effectiveness focuses on teaching processes, *teacher* effectiveness tries to identify teacher characteristics, like skills, experiences, dispositions and sometimes even personally traits, associated with teaching quality and student achievement. The following type of teacher characteristics will be briefly reviewed:

Table 3.3 Latent and manifest teaching variables based on Good et al. (2009) and Klieme (2012)

Content exposure and structure
Opportunity to learn
Curriculum alignment
Available time
Goal oriented assessment
Focus on what is important
Classroom management
Degree of student involvement
Visible and coherent planning
Structure during lessons
Providing feedback
Supportive classroom climate
Pro-active and supportive classrooms
Caring communities
Appropriate expectations
Help students to exceed
Cognitive activation
Coherent content
Sufficient depth
Thoughtful discourse
Scaffolding students' ideas and task involvement
Understanding at a higher level
Teaching meta-cognitive strategies
Authentic application of concepts in different contexts

personality traits, formal qualifications and experience, subject matter expertise and knowledge about teaching and learning and pedagogical content knowledge.

Personality Traits

Throughout the history of teacher and teaching effectiveness research personality characteristics of teachers have been investigated, looking at variables like: flexibility/rigidity, extraversion/introversion, locus of control, self-efficacy, general and verbal intelligence (cf. Brophy 1983; Darling-Hammond 1999).

As it was already referred to in a previous section reviews, carried out during the nineteen sixties and seventies, by authors like Medley and Mitzel (1963), Rosenshine and Furst (1973) and Gage (1965) showed little support for associations between personal characteristics of the teacher like being warm hearted or inflexible on the one hand, and pupil achievement on the other. In a later review, Darling-Hammond (1999), concludes that effects of general intelligence are inconsistent and small, but that some studies convincingly demonstrated a positive impact of verbal ability.

Formal Qualifications and Experience

Effects of teacher education, usually expressed in terms of formal qualifications, like having a BA or MA degree, or being certified to teach in a specific field, have traditionally been included in studies into education production functions. In developed, industrialized countries, factors like formal qualifications do not appear to make much of a difference. In developing countries, such variables appear to be more often of significant impact. The explanation for this phenomenon is probably that the variation in formal teacher training in developed countries is usually quite limited and teachers are more or less uniformly equipped to carry out their job. In developing countries teacher preparation is less uniformly distributed. One could say that in developed countries, the impact of teacher education does not come out strongly from cross sectional and comparative studies, because there is a lack of variability in the variable of interest. In Table 3.4, which combines results from two meta-analyses by Hanushek (1995, 1997), the larger impact of teacher education in developing countries is illustrated.

In a way, these results are corroborated by the outcomes of studies in the United States about alternative certification of teachers, i.e. other than official full teacher qualifications, as well as studies that have looked at out of field teaching (teaching a subject for which a teacher holds no official qualification). Wayne and Youngs (2003), when summarizing studies by Goldhaber and Brewer (1997, 2000) noted that for mathematics, results of fully certified teachers were better than for non-formally qualified or alternatively qualified teachers. Similar results were not confirmed for other subjects. In a study using state-level data from the USA, Darling-Hammond (1999), used a more fine-graded scale of teacher qualification, distinguishing between:

- teachers with full certification and a major in their field;
- teachers with full certification;
- teachers less than fully certified;
- uncertified teachers.

She found substantial positive effects for certified teachers and substantial negative effects for uncertified teachers (correlations in the order of 0.71 to −0.51).

Table 3.4 Percentages of studies with positive significant associations of resource input variables and achievement for industrialized as compared to developing countries

Input	Industrialized countries	Developing countries
	% sign. positive associations	% sign. positive associations
Teacher/pupil ratio	15	27
Teacher's education	9	55
Teacher's experience	29	35
Teacher's salary	20	30
Per pupil expenditure	27	50

Sources Hanushek (1995, 1997)

Results of studies which have investigated the effects of teacher experience are not always showing the expected positive effect. According to Darling-Hammond (1999, p. 9) effects are not always significant, nor linear. Effects of experience are particularly visible when teachers with less than 5 years of experience are included in the study.

Subject Matter Knowledge and Knowledge About Teaching and Learning

Breaking up the black box of teacher education, the most frequently addressed analytic variables in explaining why some teachers are more effective than others are subject matter mastery and pedagogical knowledge. In the more recent research literature, an interactive construct, combining the two, namely "pedagogical content knowledge" appears to show promising results.

In her review, Darling-Hammond (1999), referred to studies, which correlated teachers' course taking in subject matter areas and scores on subject matter tests to student achievement. She concludes that the former show positive effects more frequently than the latter. Low variability in test scores is seen as the main reason for low and insignificant associations. Subject matter mastery is seen as a basic requirement that is relatively uniformly addressed in initial teacher training. In this sense the explanation of the results on teachers' subject matter mastery is the same as the one already given with respect to overall teacher education effects. Hawk et al. (1985) found that the relationship between teachers' training in science and student achievement was greater in higher level science courses.

Darling-Hammond (ibid) listed some ten studies that indicate that pedagogical training generally has a stronger effect than subject matter mastery. It should be noted that most of the studies she referred to looked at subject matter related teaching methods. As suggested by Byrne and Fieldingbarnsley (1995), effects of subject matter mastery are likely to interact positively with knowledge on how to teach the subject to various kind of knowledge, meaning that the impact of subject matter mastery is augmented by subject matter related didactic knowledge. Wayne and Youngs (2003), on the other hand, presented results that showed that pedagogical training in language teaching appeared to lower student achievement.

Pedagogical Content Knowledge

In his seminal article in the Education Researcher, Lee Shulman (1986) criticized a sharp division between subject matter mastery and pedagogical skills of teachers. He introduced the concept of pedagogical content knowledge, briefly described as "subject matter knowledge for teaching". Pedagogical content knowledge is about

the selection of topics, useful forms of presentation, analogies, illustrations, examples, explanations, and demonstrations. Pedagogical content knowledge also includes understanding of what makes the learning of specific topics easy or difficult, which includes knowledge about conceptions and misconceptions that students bring to the lesson. The assumption is that "deep knowledge" about the content and structure of a subject matter area forms the crucial precondition for teachers manifesting pedagogical content knowledge in their teaching. Additional components sometimes included in the concept are knowledge on the appropriate use of teaching materials and media, as well as strategic knowledge on the application of teaching strategies. Studies investigating the effect of pedagogical content knowledge are those by Hill et al. (2008) and Baumert et al. (2010). Reviews are provided by Putnam and Borko (2000), Gess-Newsome and Lederman (1999) and Loughran et al. (2006).

Key Factors

In the previous sections, various stages in the development of instructional effectiveness have been pointed out, highlighting the main independent variables that were addressed throughout this history. The instructional and teaching variables have a central place in the input, process, output and context model that was explained in the initial section of this chapter. A reformulation of this ordering framework is given in Fig. 3.2 developed by Seidel and Steen (2005). This framework will be used to present key factors of instructional effectiveness, as well as operational variables used to measure them.

Thereby, four components are distinguished: (a) teacher professional competencies as an input measure, (b) instruction characteristics as a learning enhancing measure, (c) the quality of learning processes as a more proximal student throughput measure, (d) cognitive and non-cognitive student outcomes as more stable and long-term student measures. The components are summarised in Table 3.3 and outlined in the following section of the chapter (Tables 3.5 and 3.6).

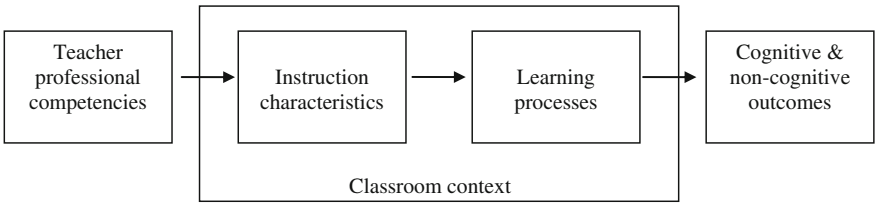

Fig. 3.2 Model of instruction effectiveness (Seidel and Steen 2005)

Table 3.5 Components of the instruction effectiveness model

Components	Description
Teacher professional competencies	Teacher characteristics that are assumed to have a direct effect in the actual teaching situation, e.g. high expectations for students, beliefs about the nature of learning
Instruction characteristics	Broad variety of instruction characteristics as reported in the research articles on instructional effectiveness
Quality of learning processes	Student throughput measures with a focus on the regulation of learning activities in the process of knowledge acquisition, e.g. students' cognitive engagement, quality of learning motivation, application of deep learning strategies
Cognitive and non-cognitive outputs	Student output measures with a focus on stable student characteristics. Cognitive outputs comprise the students' achievement, competencies, performance. Non-cognitive outputs comprise the students' interest in domains, their attitudes and beliefs as well as their self-concept of ability

Cited from Seidel and Steen (2005)

Table 3.6 Teaching variables

Instruction characteristics	Indicators
1. Learning time	Time on task, effective use of teaching time, homework, mastery learning
2. Opportunity to learn	Content covered, opportunity to learn
3. Classroom organisation	Classroom management, discipline, control
4. Orderly, functional learning environment	Learning climate, classroom climate, achievement pressure[a], Mastery-orientation, performance orientation[a]
5. Clear and structured teaching	Direct teaching, structured teaching, teacher demonstrations, teaching for basic skills, clarity
6. Activating	Cooperative learning, situated learning, discovery learning, peer-tutoring, student experiments, hands-on activities, group work, individual work, individual learning, discussions
7. Learning to use learning strategies	Cooperative learning strategy training, problem-solving, meta-cognitive training, scientific inquiry training, thinking aloud training, concept mapping, organising/structuring methods, language acquisition training, phonemic awareness training, reading strategies, writing strategies, formal learning strategy training
8. Challenge	Cognitive activation, orientation towards understanding, active student engagement, authentic contexts, relevance to students, language level, varying representation formats
9. Support (mutual respect)	Quality of teacher-student interactions, student-student interactions, teacher support
10. Feedback/monitoring	Feedback, monitoring, individual frame of reference

(continued)

Table 3.6 (continued)

Instruction characteristics	Indicators
11. Evaluation of goals/attainment	Assessments, tests
12. Adaptive teaching	Variable teaching methods, adaptive teaching, orientation towards individual learning processes, choice, taking into account student pre-requisites
13. Practice	Drill, repetitions, applications
14. Material	Quality of curriculum, textbooks, use of computers
15. "Integrated" instructional concepts	Constructivist instruction, inductive teaching, concept-oriented/integrated instruction

Cited from Seidel and Steen (2005)
[a]Reverse coded for summary

More Elaborate Conceptual Analysis of the Teaching Variables

The factors that were presented in the previous section are subsumed in a more structured way in Table 3.7. In the second column more specific information is given about the definition and measurement of the key factors. More detailed definitions are given for relatively complex constructs, such as those concerned with the main categories "activating", "learning to use learning strategies" and "challenge", as compared to concepts that have a somewhat longer tradition, e.g. structured teaching and direct teaching.

Table 3.7 More detailed definitions of core concepts

1. Learning time/opportunity to learn	
1.1 Time on task	A typical measure could be the number of on task time episodes of individual or groups of students, established in the process of direct classroom observation
1.2 Opportunity to learn	E.g. the proportion of achievement test items that, according to teachers or students, have been taught during lessons
1.3 Homework	The average time students spent on doing homework for a specific subject; the issue of whether homework will be graded enhances the relevance
1.4 Mastery learning	– Definition of expected levels of competency
	– Regular feedback
	– Corrective procedures
	– Pacing of classroom teaching process; providing additional time
1.5 Curriculum implementation	The degree to which the curriculum that is actually taught corresponds to the intended (formally planned) curriculum

(continued)

Table 3.7 (continued)

2. Classroom organization	
2.1 Classroom management	Time management by the teacher during lessons
	Preparation of teaching materials, multi media etc.
	Preparation of lesson plans
2.2 Discipline	Application of disciplinary rules, keeping order
2.3 Control	Monitoring of student behaviour during lessons
3. Orderly and functional Learning environment	
3.1 Classroom climate	Aspects of order and discipline
3.2 Achievement pressure	Students feeling urged to do well
	Graded assignments
	Reinforcement conditional on good performance
	Praise for good performance
3.3 Mastery orientation	Application of attainment targets and standards
3.4 Performance orientation	Emphasis on comparative assessment and elements of competition
4. Clear and structured teaching	
4.1 Direct teaching	1. Teaching goals are clearly formulated
	2. The course material to be followed is carefully split into learning tasks and placed in sequence
	3. The teacher explains clearly what the pupils must learn
	4. The teacher regularly asks questions to gauge what progress pupils are making and whether they have understood
	5. Pupils have ample time to practice what has been taught, with much use being made of "prompts" and feedback
	6. Skills are taught until mastery of them is automatic
	7. The teacher regularly tests the pupils and calls on the pupils to be accountable for their work
4.2 Goal directed	Teaching is led by clear goals and targets
4.3 Teacher demonstration	Teachers present material explicitly and in detail; e.g. the presentation of worked out problems in mathematics
4.4 Teaching basic skills	The procedures of goal directed, direct teaching are often applied to reading, writing and arithmetic in elementary schools; this might be translated in time for basic subjects at school, an aspect of content covered or opportunity to learn
5. Activating	
5.1 Cooperative learning	Group problem solving
5.2 Situated discovery	Within a context of discovery learning authentic material is used as contextual information
5.3 Peer tutoring	When students alternately play the role of tutor and tutee; students taught in order to be able to tutor
5.4/5.5 Individual student work	When students work individually on assignments and learning tasks

(continued)

Table 3.7 (continued)

5.6 Student discussion	When students are involved in group discussions; e.g. when triggered by open ended, controversial topics, questions with uptake
6. Learning strategies	
6.1 Cooperative learning strategies	– Perceiving the importance of group goals
	– Students being aware of the role they have to play in achieving the group goals
	– Breaking up tasks in smaller components
	– Accepting responsibilities for a sub-task
	– Sharing specific roles, e.g. record keeper
	– Teaching of interpersonal skills (listening, stating ideas clearly, providing constructive criticism, collaborative skills, equal sharing of work)
	Collaborative mental model building (Hogan 1999)
6.2 Meta-cognitive strategies	Basically two kinds of strategies
	(a) Problem solving strategies
	(b) Control strategies, like self monitoring and self regulating learning
	re (a) Systematic steps to solve problems: systematic analysis of what the problem is about, overall planning, what to do first, second, etc., linking together partial solutions, verification of hypothetical solutions, evaluation: would it have been possible to reach outcomes in a different way
	re (b) goal setting, self-observation, self-judgment, self-reaction; in short self-monitoring; control strategies involve checking what one has learned and working out what one still needs to learn, allowing learners to adapt their learning to the task at hand (OECD 2003, p. 13)
6.3 Scientific inquiry	Students apply scientific inquiry methods, by means of
	– Development of meta-cognitive skills (see above)
	– Scaffolded inquiry, reflection and generalization
	– Reflective assessment (see control strategies) (White and Frederiksen 1998)
6.4 Thinking aloud	Students talk about how they direct and reflect on tasks, while applying problem solving and control meta-cognitive strategies
6.5 Organizing methods	Organizing methods depend on confronting students with (different) ways the subject matter is ordered, e.g. a hierarchical versus a network type of organization of subject matter elements; students learning processes are steered by means of questions, while their progress is monitored and feedback is required, in order to stimulate self-regulated learning (Chularut and DeBacker 2004; Einsiedler and Treinies 1997)
6.6 Language acquisition	Organizing methods, specific for mother tongue and foreign language learning

(continued)

Table 3.7 (continued)

6.7 Reading/writing strategy training	Idem, see above, organizing methods specific for reading and writing tasks, e.g. applying morphemic and contextual analysis, as alternative strategies in vocabulary instruction; or the development of phonemic awareness (Byrne and Fieldingbarnsley 1995)
	In summary, learning to learn domain specific processing depends on two components: the (deep) structure of the subject matter area, and the application of meta-cognitive strategies (problem solving as well as control strategies)
6.8 General strategies	See 6.2
7. Providing a challenging learning environment	
7.1 Cognitive activation	The stimulating of understanding by means of cognitive demands and the attempt to get at reasoning behind an idea, and the "deep" structures of processing. More concrete steps
	– Use of higher-order questions to challenge students in their learning
	– Working on problems for which the solution is not immediately clear
	– Stimulate insightful learning on the basis of situated and contextualized problem environments
	– Learning from mistakes
	– Self-evaluation and self-reflection
7.2 Active student engagement	Cognitive activation combined with group work
7.3 Authentic contexts, relevance	Yair (2000) Instructional motivation theories, which postulate that students' learning experiences are optimized when instruction is authentic, challenging, demands skills and allows for student autonomy. Authenticity is defined as: the extent to which an activity is important for the students' immediate and long-term aims
7.4 Language level	The idea that language should be at appropriate but also high cognitive level
7.5 Representation formats	Inviting and varied representation formats as means to evoke student engagement
8. Support	
8.1 Quality of interactions and teacher support	The following aspects of *classroom climate*
	Positive relationships between students
	Positive relationships between students and teachers
	Students feeling supported
	Hill and Rowe (1998) mention the following aspects of teacher support
	– Teacher empathy
	– Energy and enthusiasm
	– Fairness and firmness
	– Helpfulness and responsiveness
	– High expectations

(continued)

Table 3.7 (continued)

	– Quality of instruction
	– Appropriateness of instruction
	– Feedback
	– Sufficient time for accomplishing learning tasks
9. Feedback	
9.1 Feedback, monitoring	Characteristics of feedback given in the context of mastery learning are that feedback is based on comparing current achievement against a "mastery" level of achievement that such feedback is given rapidly and intended to be diagnostic, and, finally, that students are given the opportunity to discuss with their peers how to remedy any weaknesses
	Instrumental feedback works better than judgmental feedback (instrumental feedback provides task related information that allows to improve future performance)
10. Evaluation	
10.1 Assessments/tests	Formative assessment, frequent monitoring of students' progress
11. Teacher characteristics	
11.1 High expectations	Teachers hold high expectations about students' capacities. This variable has really two components
	– Achievement press
	– A positive, "optimistic", "non-defeatist" attitude of students' capacities
11.2 Constructivist beliefs about learning	*Constructivist-Compatible Instruction* is based on a theory of learning that suggests that understanding arises only through prolonged engagement of the learner in relating new ideas and explanations to the learner's own prior beliefs. A corollary of that assertion is that the capacity to employ procedural knowledge (skills) comes only from experience in working with concrete problems that provide experience in deciding how and when to call upon each of a diverse set of skills
	Traditional Transmission Instruction is based on a theory of learning that suggests that students will learn facts, concepts and understandings by absorbing the content of their teachers' explanations from a text and answering related questions. Skills (procedural knowledge) are mastered through guided and repetitive practice of each skill in sequence, in a systematic and highly prescribed fashion, and done largely independent of complex applications in which those skills might play some role (Ravitz et al. 2000, p. 3)
	Teacher manifesting a constructivist belief on teaching would be likely to score high on items like:
	– Students are capable of finding their own solutions to problems

(continued)

Table 3.7 (continued)

	– Students should be given ample opportunities for independent work
	– Relationships between facts should be discovered by students
	– The teacher is seen as a facilitator of independent learning
12. Adaptive Teaching	
12.1 Various teaching methods	Differentiation in teaching methods depending on student aptitudes
12.2 Adaptive teaching	Differentiation in didactic approaches depending on student aptitudes
12.3 Open tasks, choice	Students work independently on tasks in which they can select difficulty levels that fit their current state of knowledge and skills
12.4 Student pre-requisites	Differentiation of learning tasks contingent on current achievement level as determined by adaptive tests
13. Student practice	
13.1 Drill/repetition	Prolonged practice and repetition of mastering subject matter and skills
13.2 Application	Students getting ample practice to apply skills and knowledge to new situations or related problems
14. Material	
14.1 Textbooks	Quality of textbooks
14.2 Media	Appropriate use of media, computers included
15. Integrative approaches	
15.1 Constructivist approach	See 11.2
15.2 Inductive approach	Approach in which students have to discover regularities and relationships between single facts and events
15.3 Concept-oriented approach	E.g. concept mapping in foreign language instruction, which is a tool for representing the interrelationships among concepts in an integrated hierarchical way

Cited from Scheerens et al. (2007)

A More Synthetic Use of the Factors

Since the number of main categories (15) is relatively large, it is tempting to find more general categories, to be more concise. The following major dimensions could be taken into consideration:

– a curricular dimension, containing, opportunity to learn, strategies to learn about the deep structure of domain specific knowledge, and textbooks;
– a dimension of teacher orchestrated classroom management and climate creation, including time, achievement orientation, high expectations, disciplinary climate, activating measures, such as variation in representation formats, media,

forms of practice, variation in applications (theoretical and authentic) grouping forms and differentiation/adaptive teaching;
- a teaching strategy dimension with two main sub-categories:

  (a) structured, direct teaching, mastery learning orientation, drill and practice;
  (b) constructivist-oriented teaching strategy, teaching meta-cognitive strategies, cognitive activation, frequent open learning tasks, discovery learning, fading from more structured to more open assignments

- a climate dimension, support and positive interactions;
- a dimension representing evaluation and feedback

Both the original set of 15 factors the one reduced to six factors was used for the meta-analyses, which will be presented in Part 2 of his book (Chap. 8).

References

Anderson, L. W. (1991). *Increasing teacher effectiveness*. Paris: UNESCO, IIEP.

Anderson, L. W. (2004). *Increasing teacher effectiveness* (2nd ed.). Paris: IIEP, UNESCO.

Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. OECD. (2003). *Learners for life. Student approaches to learning. Results from PISA 2000.* Paris: OECD publishing.

Bangert, R. L., Kulik, J. A., & Kulik, C. C. (1983). Individualized systems of instruction in secondary schools. *Review of Educational Research, 53*, 143–158.

Baumert, J., Blum, W., & Neubrand, M. (2001). *Surveying the instructional conditions and domain-specific individual prerequisites for the development of mathematical competencies*. Paper presented at the special meeting on the co-ordination of national teacher components for PISA 2003, May 28, 2001, Munich, Germany.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 133–180.

Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. H. (2014). *International perspectives on teacher knowledge, beliefs and opportunities to learn. Teachers education and development study in mathematics (TEDS-M)*. Dordtrecht: Springer.

Bloom, B. (1976). *Human characteristics and school learning*. New York: McGraw Hill.

Brophy, J. (1983). Conceptualizing student motivation. *Educational Psychologist, 18*(3), 200–215.

Brophy, J. (2001) Generic aspects of effective teaching. In: M. C. Wang & H. J. Walberg (Eds.), *Tomorrow's teachers*. McCutchan Publishing Company.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In: M. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 328–375). New York: MacMillan.

Byrne, B., & Fieldingbarnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young-children—a 2-year and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*(3), 488–503.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 722–733.

Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*(3), 248–263.

Cohen, D. K. (1988). Teaching practice … Plus ça change … In: Ph. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice*. Berkeley, CA: McCutchan.

Collins, A., & Stevens, A. (1982). Goals and strategies of inquiry teachers. In: R. Glaser (Ed.), *Advances in instructional psychology*, Vol. II. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cotton, K. (1995). *Effective schooling practices: A research synthesis*. 1995 Update. School Improvement Research Series. Northwest Regional Educational Laboratory.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London, New York: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2012). *Improving quality in education. Dynamic approaches to school improvement*. London, New York: Routledge.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Davies, J. K. (1972). Style and effectiveness in education and training: A model for organizing, teaching and learning. *Instructional Science*, *2*.

de Klerk, L. F. W. (1985). ATI-onderzoek en differentiatie: een reactie. *Pedagogische Studiën, 62*, 372–375.

Dougherty, K. (1981). After the fall: Research on school effects since the Coleman report. *Harvard Educational Review, 51*, 301–308.

Doyle, W. (1985). Effective secondary classroom practices. In: M. J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook*. Washington: U.S. Government Printing Office.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–27.

Einsiedler, W., & Treinies, G. (1997). Effects of teaching methods, class effects, and patterns of cognitive teacher-pupil interactions in an experimental study in primary school classes. *School Effectiveness and School Improvement, 8*(3), 327–353.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *Special Issue of the International Journal of Educational Research*, *11*(2).

Gage, N. (1965). Desirable behaviors of teachers. *Urban Education, 1*, 85–95.

Gagne, R. (1971). *The conditions of learning*. New York: Holt, Rinehart & Winston.

Gess-Newsome, J., & Lederman N. G. (Eds.). (1999). *Examining pedagogical content knowledge*. Dordrecht: Kluwer.

Goldhaber, D. D., & Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in school finance, 1996* (pp. 197–210). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129–145.

Good, T. L., & Brophy, J. (1986). School effects. In: M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: McMillan Inc.

Good, T. L., Wiley, C. R. H., & Florez, I. R. (2009). Effective teaching: An emerging synthesis. In G. Dworkin (Ed.), *International handbook of research on teachers and teaching* (pp. 803–816). New York: Springer.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *The World Bank Research Observer, 10*, 227–246.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*, 141–164.

Hattie, J. (2009). *Visible learning*. Abingdon: Routledge.

Hattie, J., & Alderman, E. M. (Eds.). (2012). *International guide to student achievement.* London, New York: Routledge.

Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education*, *36*(3), 13–15.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education., 39*(4), 372–400.

Hill, P. W., & Rowe, K. J. (1998). Modelling student progress in studies of educational effectiveness. *School Effectiveness and School Improvement, 9*(3), 310–333.

Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching, 36*(10), 1085–1109.

Hopkins, D., Stringfield, S., Harries, A., Stoll, L., & Mackay, T. (2014). School and system improvement. *School Effectiveness and School Improvement, 25*, 257–281.

Klieme, E. (2012). Qualities and effects of teaching. Towards a conceptual theory of teaching. Key note address. EARLI SIG meeting, Educational Effectiveness, Zuerich, August 23.

Klieme, E., Rakoczy, K. (2003). Unterrichtsqualitaet aus Schuelerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhaenge mit Effeckten von Unterricht. In: Baumert, J., Artelt, C., Klieme E., Neubrand, J., Prenzel, M., Schiefelt, U., Schneider, W., Tillmann, K. *PISA 2000. Ein differenzierter Blick auf die Laender der Budesrepublik Deutschland.* Opladen: Lekse+Budrich.

Kulik, C. L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of research findings. *American Educational Research Journal, 19*, 415–428.

Kyle, M. J. (1985). Reaching for excellence: An effective schools sourcebook. Washington DC: US Government Printing Office.

Levine, D. K., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, Wise: National Centre for Effective Schools Research and Development.

Lortie, D. C. (1973). Observations on teaching as work. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.

Loughran, J., Berry, A., & Mullhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam, The Netherlands: Sense Publishers.

McBerr, H. (2000). Research into teacher effectiveness: A model of teacher effectiveness (Research report, No. 216). Norwich: The Crown Copyright Unit.

Medley, D., & Mitzel, H. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Muijs, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H., & Earl, L. (2014). Teaching effectiveness. A state of the art review. *School Effectiveness and School Improvement, 25*, 231–257.

Muijs, D., & Reynolds, D. (2001). *Effective teaching*. London: Paul Chapman.

Murnane, R. J. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record, 83*, 19–35.

National Center for Education Statistics (NCS). (2002). *Fundamental problems in the measurement of instructional processes*. NCS: Washington D.C.

Neufeld, E., Farrar, E., & Miles, M. B. (1983). *A review of effective schools research: The message for secondary schools*. Washington D.C: National Commission on Excellence in Education.

Postlethwaite, T. N., & Ross, K. N. (1992). *Effective schools in Reading. Implications for educational planners*. The Hague: IEA.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 1–4.

Ralph, J. H., & Fennessey, J. (1983). Science or reform: some questions about the effective schools model. *Phi Delta Kappan, 64*(10), 689–695.

Ravitz, J., Becker, J., & Wong, Y. T. (2000). Constructivist-compatible beliefs and practices among U.S. teachers. Center for Research on Information Technology and Organizations University of California, Irvine and University of Minnesota.

Reezigt, G. J. (1993). *Effecten van differentiatie op de basisschool (Effects of grouping in primary education)*. Groningen: RION.

Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement, 4*(1), 37–58.

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., Teddlie, C., et al. (2014). Educational effectiveness research (EER): A state of the art review. *School Effectiveness and School Improvement*, *25* 197–230.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.

Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development, 54*(1), 1–29.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.

Scheerens, J. (2000). *School effectiveness in developed and developing countries; a review of the research evidence*. World Bank paper. Available on http://www.worldbank.org/education/schools.

Scheerens, J. (2014). School, teaching, and system effectiveness: some comments on three state-of-the-art reviews. *School Effectiveness and School Improvement, 25*(2), 282–290.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organization and Management.

Scheerens, J., Vermeulen, C. J. A. J., & Pelgrum, W. J. (1989). Generalizibility of instructional and school effectiveness indicators across nations. *International Journal of Educational Resesarch, 13*(7), 789–799.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research 77*(4), 454–499.

Seidel, T., & Steen, R. (2005). The indicators on teaching and learning compared to the review of recent research articles on school and instructional effectiveness. In: J. Scheerens, T. Seidel, B. Witziers, M. Hendriks, & G. Doornekamp. (2005). *Positioning the supervision frameworks for primary and secondary education of the Dutch Educational Inspectorate in current educational discourse and validating core indicators against the knowledge base of educational effectiveness research*. Enschede/Kiel: University of Twente/Institute for Science Education (IPN).

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*, 4–14.

Slavin, R. E. (1987). *Cooperative learning: Student teams* (2nd ed.). Washington D.C.: National Education Association Professional Library.

Slavin, R. E. (1996). *Success for all*. Lisse: Swets & Zeitlinger.

Stallings, J. (1985). Effective elementary classroom practices. In: M. J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook*. Washington, D.C.: US Government Printing Office.

Stallings, J., & Mohlman, G. (1981). *School policy, leadership style, teacher change and student behavior in eight schools*. Final report to the National Institute of Education, Washington D.C.

Sweeney, J. (1982). Research synthesis on effective school leadership. *Educational Leadership, 39*, 346–352.

Van der Hoeven-Van Doornum, A. A. (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen [The impact of teachers' expectations on educational careers]*. Nijmegen: ITS/OOMO.

Van Laarhoven, P., & De Vries, A. M. (1987). Effecten van de interklassikale groeperingsvorm in het voortgezet onderwijs. Resultaten van een literatuurstudie. (Effects of heterogeneous

grouping in secondary schools). In: J. Scheerens & W.G.R. Stoel (red.), Effectiviteit van onderwijsorganisaties. Lisse: Swets & Zeitlinger.

Vedder, P. H. (1985). *Cooperative learning. A study on processes and effects of cooperation between primary school children*. The Hague: SVO.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89–122.

Weeda, W. C. (1986). Effectiviteitsonderzoek van scholen. In: J. C. van der Wolf & J. J. Hox (red.), *Kwaliteit van het onderwijs in het geding* [About education quality]. Publicaties van het Amsterdams Pedogogische Centrum, nr. 2. Lisse: Swets & Zeitlinger.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

Yair, G. (2000). Reforming motivation: How the structure of instruction affects students' learning experiences. *British Educational Research Journal, 26*(2), 191–210.

Chapter 4
Modelling School Effectiveness

Keywords School effectiveness · School leadership · School management · School effectiveness research · Professional bureaucracy · The learning organization · Schools as high reliability organizations · Participatory leadership · Single and double loop learning · School climate · School culture

Introduction

Within the framework of multi-level educational systems, the school level should be seen from the perspective of creating, facilitating and stimulating conditions for effective instruction at classroom level. As major vehicles of support to teaching could be seen:

- the establishment of action plans such as school curricula, mission statements and preferred teaching strategies;
- coordination of goals, structural arrangements and the primary process of teaching, among staff members;
- school policies and practices concerning evaluation, assessment and monitoring;
- provision of adequate teaching facilities, such as text books, computers, multimedia applications;
- external contacts in the service of real-life applications and authentic learning environments;
- facilities for professional development of teachers and in-service training;
- engagement of parents in facilitating teaching.

In addition, more general organizational and management conditions could be distinguished that do not impinge directly on teaching, but help in creating favourable organizational conditions, such as a safe, achievement oriented and pleasant working climate, external contacts and clarity and stability of structural and procedural arrangements.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_4

Finally, like in the case of classrooms, one could discuss the ecology of the school, in terms of characteristics of the composition of students, the composition of the staff, the local neighbourhood and school size.

Results of Empirical School Effectiveness Research: A First Overview of Factors

In this chapter, the focus is on conceptual and theoretical organization models that are applicable to schools. As will be further documented in subsequent chapters, the developmental history of educational and school effectiveness research has not been particularly theory driven and shows rather a more inductive, empiricist approach. In order to show the connection of the conceptual models with research practice, this paragraph gives a brief overview of the key factors addressed in school effectiveness research.

Most of the school organizational factors mentioned in the introduction have been addressed in empirical school effectiveness research. Scheerens et al. (2003, Chap. 3, Table 7), provide an overview of reviews carried out in the nineties. Main effectiveness enhancing variables mentioned in review by Purkey and Smith (1983), Levine and Lezotte (1990), Scheerens (1992), Cotton (1995) and Sammons et al. (1995) are: school climate (achievement-oriented policy, a cooperative atmosphere and an orderly climate), clear goals concerning basic skills, frequent evaluation, staff professional development, strong leadership and time on task. Other variables mentioned, not in all, but in subsets of these review studies are opportunity to learn, consensus, and cohesion among staff and parental involvement. This set of factors was basically re-confirmed in an overview by Scheerens et al. (2003, Chap. 12), in which a further specification of each factor into specific components was added (more details are provided in the next chapter).

In a recent review study (Reynolds et al. 2014), summarized by Scheerens (2014), the following list of key variables emerged (Table 4.1).

Table 4.1 Effectiveness enhancing conditions referred to in the review study by Reynolds et al. (2014, as summarized by Scheerens (2014)

Educational Effectiveness Research
Effective leadership
Academic focus
A positive orderly climate
High expectations
Monitoring progress
Parental involvement
Effective teaching (time)
Staff professional development
Pupil involvement

Again there is almost perfect overlap with the reviews carried out in the nineteen nineties and in 2003. It can therefore be concluded that there appears to be a lot of consensus in the history of educational effectiveness research on the kind of school-level conditions that matter. In the next sections, more theoretical models and theories of the school will be considered as a first basis for a deeper understanding of effectiveness enhancing school conditions.

Models and Theories of the School

Fend's Basic Theory of the School

In his earlier "Theorie der Schule" (Fend 1981) and his recent "Neue Theorie der Schule" (Fend 2006), Helmut Fend presents the school and its societal functions. In the new theory of the school, he sketches the school as part of a multi-level education system that is governed by institutional norms and formal regulations on the one hand, and by dynamic interactions that create a certain leeway for self-development and self-reference at each level. The two major schools of thoughts behind this analysis are institutionalism (perhaps "new" institutionalism) and system dynamics (particularly inspired by the work of the German sociologist Luhmann).

Both dimensions, formal regulation and dynamic acting in a social context are represented in "Plan" and "Praxis" as in Table 4.2, cited from Fend (2006, p. 177).

Fend's basic theory of the school should be seen as a foundational description, while subsequent theoretical interpretations of the school put alternating emphasis on either the formal institutional dimension or the "Praxis" dimension. To the degree that national educational systems devolve authority to the school level, and schools are autonomous in several functional domains (such as the curriculum, the assessment, finance and personnel policy), formal regulations at higher levels are less prominent, and schools have to develop their own identity to a larger degree.

The Professional Bureaucracy: Professional Autonomy

The concept of the professional bureaucracy was developed by Mintzberg (1979). The main characteristics of the school as a professional bureaucracy are the following ones:

- the internal cohesion of the organization depends predominantly on the standardization of skills of the functionaries—teachers in our case—which is based on long-specialized training;
- a large degree of professional autonomy of the teachers, whereby loyalty towards the organization has to compete with loyalty to the profession and loyalty to the "client";

Table 4.2 Fend's theory of the school

Plan (*Gestaltungsformen*)	Praxis (*Realitaeten*)
Society formation: legal context	**Actual political interrelationships**
The legal system of a society	The reality of public policy making and political practices at different levels
The formal domain of education	**Actual relational structure at school**
School laws	Realized subject matter offerings
Curricula	Teaching methods and study books
Learning routes	Selection processes
Tests	The way marks are given
Formal determination of work at school	**Actual acting of teachers**
Rules	Actual teaching practice
Formal rights and duties	Matching of school rules and classroom culture
Formal authority	Style of leading the class
Formal duties of the students	**Actual learning of students**
Presence	Acquiring skills or "sitting out" the school period
Discipline	Relationships with peers
Tasks and homework	Motivated or unmotivated confrontations with subject matter

- a relatively underdeveloped interest in the external environment; the basic assumption in the professional bureaucracy is that the environment may be complex but is, at the same time, stable;
- a specific role for leadership and management, which is seen as mostly administrative and not substantive in the sense that school leaders are expected to give direction to teachers; but rather play a submissive and supportive role;
- technology in the professional bureaucracy has on the one hand the nature of a "well-stocked tool box", but on the other hand, holds the challenge of adapting these standard tools and solutions to ever changing circumstances in the work with clients (in this case pupils);
- there is little readiness and openness for change and opposition against rationalization of the work and monitoring performance among the professionals;
- recruitment of personnel is the most important control measure within the organization; within the framework of the profession as such, adaptation of the initial training is the most important control mechanism.

The concept of the school as a professional bureaucracy is related to Weick's image of the school as a "loosely coupled organization". (Weick 1976) "Loose coupling", according to Weick, refers to a relatively small interdepence among subsystems like teachers among themselves, and head teachers and teachers. At the same time, there is also little cohesion between aspects of the organization's functioning, like the coupling of means and goals and between decisions planned

and actual implementation. An example of this last phenomenon is the well-known situation (at least in the Netherlands) where schools used to have a nicely phrased "school work plan", which was safely put away in a cupboard and had little relationship to what was actually happening (cf. Van der Werf 1988). As far as technology is concerned, Weick emphasizes the "fuzzy" technology of schools, where there is little consensus about goals and means and evaluation of central means-to-end relationships is difficult.

What kind of school management would fit in an organizational structure as depicted in the images of the professional bureaucracy and the loosely coupled system? The general answer would seem to be that such structures require only minimal management. In such a structure, little need is felt for long-term planning and strategy development. In the prototype form, there would be no intermediary structures and hence no middle-management. Operational management is firmly in the hands of the professionals (teachers) in the operating core (the classroom) of the organization. The metaphor of the teacher as the King/Queen in his/her classroom comes to mind. Monitoring and performance control will tend to be seen as threats to the professional autonomy. According to this theoretical image, the most potent management domain, which is human resource management, is in the actual practice of most countries, strongly limited because of fixed conditions of labour.

In short, the image of a professional bureaucracy is explicit in warning us for the limitations of trying to develop a type of management in schools, which touches the primary process of teaching and learning. Interestingly this orientation is exactly the central focus in the concept of "educational" or instructional leadership, as developed in the context of effective schools research.

Prescriptive Interpretations of the "Effective Schools Model": The Rationally Managed School

Empirical school effectiveness research basically addresses the question which organizational and instructional conditions explain why some schools have better results, in terms of student achievement than others, after taking differences in the student intake between schools into account. A summary of school characteristics that have been identified by this research strand has already been given in a previous section. When taken together, this set of effectiveness enhancing school characteristics provides an image of a more managed school that even resembles a business firm with more pronounced leadership, something like a mission, outcome orientation, and internal cohesion and coordination. The leadership concept associated with the effective schools model is instructional or educational leadership, which is a concept that fits very well with the basic query about across level linking mechanisms in the analysis of educational systems as hierarchical loosely coupled systems (compare Chap. 1). This is so because the central idea of instructional leadership is the facilitation of the primary process of teaching and learning in schools. In this section, a closer look will be taken at the concept of instructional or

educational leadership (the two terms will be used intermitted, as they are taken to have the same meaning).

In the operational definitions and instruments concerning educational leadership, a general division into two conceptions can be made:

(a) general leadership skills applied to educational organizations:

- articulated leadership,
- information provision,
- orchestration of participative decision making,
- coordination.

(b) instructional/educational leadership in a narrower sense, i.e. leadership directed at the school's primary process and its immediate facilitative conditions:

- time devoted to educational versus administrative tasks,
- the head teacher as a meta-controller of classroom processes,
- the head teacher as a quality controller of classroom teachers,
- the head teacher as a facilitator of work-oriented teams,
- the head teacher as an initiator and facilitator of staff professionalization.

Of these two dimensions, the second, namely leadership focused on the school's primary process, should be considered as central. The other dimension addresses the specific demands required for leading and controlling organizations in which professionals at the operating core need to have a considerable degree of autonomy.

As a whole educational leadership can be seen as a phenomenon that needs to strike a balance between several extremes: direction versus giving leeway to autonomous professionals, monitoring versus counselling and using structures and procedures versus creating a shared (achievement oriented) culture. Sammons et al. (1995) in this context refer to the school leader as the leading professional.

The system-theoretical concept of meta-control is perhaps the most suitable to express the indirect control and influence an educationally or instructionally oriented school leader exercises on the school's primary process. Of course this does not imply that the head teacher is looking over the teachers' shoulder all the time, but he or she is 'involved' in the important decisions on objectives and methods, and visibly cares about overall achievement levels and individual pupils' progress. From the set of components that were listed above it is evident that the meta-control of the school leader is exercised in a non-authoritarian way, expressing concern about pupils, individual staff members, and team work.

Some authors, who define educational leadership, say more about structural conditions surrounding the instructional process, whereas others are more focused on cultural aspects. Irwin (1986, p. 126) belongs to the former category in mentioning the following aspects of educational leadership:

the school leader:

– functions as an initiator and coordinator of the improvement of the instructional program;
– states a clear mission of the school;

- has a task-oriented attitude;
- establishes clear objectives;
- supports innovation strategies;
- stimulates effective instruction;
- is quite visible in the organization;
- sees to it that pupils' progress is monitored regularly;
- delegates routine tasks to others;
- regularly observes both the work of teachers and pupils.

Leithwood and Montgomery (1982, p. 334) mention the following more cultural aspects of educational leadership:

- stimulation of an achievement-oriented school policy;
- commitment to all types of educational decisions in the school;
- stimulating cooperative relationship between teachers, in order to realize a joint commitment to the achievement-oriented school mission;
- advertising the central mission of the school and obtaining of support of external stakeholders.

A central dimension in concepts of leadership at school is the division of responsibility across functional domains of the school between leaders, deputy leaders and teachers.

Scheerens (2012) describes the conceptual development and phases of model emphases as "the full circle of leadership distribution":

"The classical model of the school, the professional bureaucracy, places much autonomy with the teachers. In professional bureaucracies training of teachers is the most important coordination mechanism. Based on their professional skills teachers can operate quite autonomously and there is little need for direct supervision. According to Weick (2001) professionals operating in loosely coupled structures are well equipped to react adaptively to small changes in the environment (e.g. changes in the composition of the student population). A professional bureaucracy can exist without pronounced leadership. The school effectiveness and school improvement movements expect teachers to become involved in cooperative activities, to stimulate common orientations on the ends and perhaps also the means of schooling. Strong instructional leadership was the boldest jump forward in breaking the traditional structures with goal coordination (an achievement-oriented school culture and emphasis on basic skills) as the main orientation, next to creating an orderly atmosphere and stressing the monitoring of students' progress. Transformational leadership, more of an invention of school improvement experts, than of school effectiveness researchers, retreated somewhat from the bold intrusion in the direction of the primary process of teaching and learning and concentrated on organizational development and organizational learning with a looser or in any case much more indirect relationship to enhancing student outcomes. One could read the move towards distributive or collective leadership, sometimes also described as "teacher leadership", as a kind of re-installation of the professional autonomy of teachers. In some studies leadership (defined as distributed leadership) and the

Fig. 4.1 The full circle of concept development on school leadership; Scheerens (2012)

organizational development that it is supposed to stimulate (more collaboration) appear almost one and the same. When Hallinger and Heck (2010) speak of organizational leadership; the idea of individualized or focused leadership has fused with "structural coordination", in the sense of structural and sustained procedural arrangements, such as cooperative structures, task divisions and common standards. This concept seems to come close to the much older idea of "substitutes for leadership", coined by Kerr and Jermier (1978)". The full circle of leadership distribution is depicted in Fig. 4.1 reproduced here.

"This circular movement, however, should not be read as a return to square one. Under the influence of more turbulent environmental conditions a modernization process has undoubtedly occurred in which schools have become more integrated structures and pay more attention to secondary processes (coordinating, evaluating, maintaining external contacts). This means that alignment might gradually replace loose coupling, and that collegial support, data driven teaching and task-related cooperation between teachers are considered as important levers of instructional improvement (see the seminal review by Elmore 2000). At the same time the development in the conceptualization of leadership can be read as a gradually return to the recognition of the importance of teachers as resourceful practitioners. The question might be raised to what extent this process has led to an accompanying retreat of personal leadership ambitions. The leadership functions are still in place, and have sometimes even been elaborated. Personal leadership still has a role in school organizations that are on the one hand more integrated, but on the other still maintain important characteristics of "loose coupling". This role might be well captured with the concept of "meta-control", which literary means control of control. School leadership as meta-control would seek to fully exploit the potential of distributed leadership, organizational leadership and substitutes for leadership while maintaining an overarching outlook on the well-functioning of the whole" (ibid., p. 133–134).

The introduction of indirect effect models in school leadership effect studies invites reflection on hypothetical causal associations between leadership variables, intermediary school factors and educational outcomes. Here a connection could be made with integrative or comprehensive, multi-level educational effectiveness models as introduced in Chap. 1. Also in: Scheerens (1992), Creemers (1994), Stringfield and Slavin (1992), Creemers and Kyriakides (2008), Huber and Muijs (2010). Generally speaking these models try to define nested structures of facilitating conditions, teachers facilitating student learning and school organizations and school leaders facilitating effective instruction. In actual fact, the connection to comprehensive educational effectiveness models has not been made in the major empirical leadership studies, until recently; e.g. Bruggencate et al. (2013), Luyten (2009) and Heck and Moriyama (2010). An overview of indirect effect studies is given in Scheerens (2012). Scheerens (2012, ibid.) combines leadership trait, style and behavioural characteristics in a comprehensive conceptual map of leadership and potential intermediary school factors.

Figure 4.2, from Scheerens (2012), shows a global division between person and task-related strategies. So far leadership studies have particularly focused at

Relevant personality traits and competencies	**Leadership style**	**Leadership behaviour**	**Effectiveness enhancing factors**
Extraversion social appraisal skills		External contacts Buffering	Enhanced teaching time
intelligence motivation internal locus of control domain specific knowledge conscientiousness	Task-related →	Direction setting (goals, standards **Monitors curriculum and instruction (managing the instructional program**)	Clear goals and standards Opportunity to learn Student monitoring & feedback Structured teaching Active teaching Active learning
Extraversion Social appraisal skills Self confidence	Person-related→	HRM & HRD Coaches teachers Recruits teachers Builds consensus	Cohesion among teachers Professionalization Teacher competency Teachers' sense of self efficacy
Basic human values General moral beliefs Role responsibility		Sets values Creates climate	Shared sense of purpose among teachers High expectations Disciplinary climate Supportive climate

Fig. 4.2 Intermediary causal structure of leadership at school

person-oriented approaches, building consensus and cooperation. A new emphasis on facilitating teaching and learning, not only by people strategies but also by means of "technology" in the broadest sense, might become an additional emphasis in future work.

The School as a Learning Organization: The Cybernetic Principle[1]

Organizational Learning can be defined in three different ways:

(a) as the sum total of individual learning of the members of the organization
(b) in the sense of enhancing the organization's instrumental effectiveness (single-loop learning)
(c) in the sense of enhancing the organization's external responsiveness (double-loop learning)

Re (a) *individual learning*. Particularly when organizations are knowledge intensive, as is the case with educational organizations, there is the strong expectation that workers will keep their knowledge and skills "up-to date". In the corporate, world rapidly changing technology and markets are the basic motives for training and human resource development (hrd) activities. In this setting, there is a growing interest in a conception of hrd that depends less on formal training, but situates learning in the working place itself, in "learning by doing", subsequently integrating training responsibilities in management functions throughout the organization. Of course something "extra" is required to convert individual learning into organizational learning. All coordination mechanisms that are known from the organization literature can play a role in orchestrating individual learning in a way that the benefits for the organization as a whole are maximized. Examples are: a clear mission and result orientation of the organization, organizational structures that enable exchange between units and subunits, facilitation and supporting technology (i.e. "group-ware") for communication and collaboration between members of the organization and even standardization of outcomes and processes. This latter coordination mechanism does not fit in so well with the expectations of flexibility and a more "organic" functioning of "learning organisations", however.

Re (b) *organizational learning as single—loop learning*. The concepts of single- and double-loop learning, as introduced by Argyris and Schön (1974) form the core of the theoretical basis of learning organizations. Single-loop learning rests in an ability to detect and correct error in relations to a given set of operating norms (Morgan 1986, 88). In its turn single-loop learning should be seen against the conceptual background of cybernetics ("steermanship"), which sees the self-regulation of organisms and organizations as based on processes of information

[1]This section is based on Scheerens et al. (2003), Chap. 5.

exchange involving negative feedback. Learning in this sense is characterized as a gradual shaping of behaviour, constantly correcting for mistakes or suboptimal solutions. In social contexts, "right" and "wrong" are determined by agreements and norms, hence the qualification of the kinds of norms that are central in single-loop learning. When these are defined as the operating norms, they should be taken as the preferred end-states of an organization's primary process, or the objectives of the organization's core business. Single-loop learning takes these objectives as given and concentrates on the optimal selection of means and technology to attain these objectives. This instrumental perspective is quite similar to the approach of school effectiveness research, in which scientific methods are used to find out which organizational and instructional conditions are most effective in realizing key-outcomes. In a less stylised from the day to day running of an organization can also be seen as guided by this instrumental approach. In case of organizations with "unclear" technologies, such as schools, such a trial-and-error approach to improving the effectiveness of the primary process appears quite relevant, at least in theory. In actual practice, such organizations are also likely to have quite a few barriers that work against a learning orientation (see the earlier section on schools as professional bureaucracies). Single-loop learning emphasizes the need for information that can shape a gradual improvement of primary and supporting organizational processes in obtaining basic outcomes.

Re (c) double-loop learning. "Double-loop learning depends on being able to take a "double look" at the situation by questioning the relevance of the operating norms" (Morgan 1986, 88). So, double-loop learning does not take prefixed operating norms (or objectives) for granted, but makes them the object of analysis and reflection. The basic motive to choose this approach is grounded in an open-systems view of organizations, where situational conditions set the stage for defining what organizational effectiveness means. Contingency theory has provided further insight in the kind of situational conditions that matter: changes in the predictability of the environment and the nature of the organization's technology being the most prominent types of "contingency factors". The more dramatic the dynamics of these situational conditions, the stronger the need for critical review of the organization's operating norms and "double-loop learning". The type of analysis and information gathering that is required for double-loop learning cannot stop at an internal review of "instrumental effectiveness", but also needs an external scan of situational conditions. The emphasis on monitoring with an open-minded about operating norms and objectives resembles the orientation of "backward evaluation" and "retroactive planning" described in the previous section. Analysis of the organizational structures that facilitate or hinder organizational learning in the sense of double-loop learning form the basis of further clarification of the concept of the learning organization. Before doing so, it is important to realize that the relevance of this concept, particularly as far as double-loop learning is concerned, strongly depends on the dynamics of situational factors.

Morgan (ibid., 89.90) mentions three types of failures of organizations in implementing double-loop learning.

First, formal planning approaches including organizational goals and objectives, clearly defined roles and bureaucratic structure with pronounced hierarchy, create fragmented structures "that do not encourage employees to think for themselves". Fragmented operation of the organization is further seen as to be encouraged by political processes in which each subunit pursues its own goals and means are treated more or less as ends in themselves (ibid., 89). It is interesting to note that the author judges highly sophisticated single-loop learning systems in such bureaucratic contexts as actually preventing double-loop learning, "since people are unable or not prepared to challenge underlying assumptions" (90).

Bureaucratic accountability systems, where people are held responsible for their performance within a system that rewards success and punishes failure, is seen by Morgan as a second barrier to double-loop learning. He sees such systems as fostering defensiveness of employees and as an incentive for covering up and "impression management" (make situations look better than they actually are). He also criticizes the tendency to oversimplification as complex issues are difficult to address in such a context.

The third barrier to double-loop learning, mentioned by Morgan is the tendency of organizations to rationalize and meet problems with rhetoric. Organizations develop "theories in use" that may be socially reinforced to constructions that are insufficiently rooted in reality.

According to Morgan, these barriers can be overcome by encouraging openness and reflectivity and a divergent thinking approach to the analysis and solution of complex problems, which means that the importance of exploring different viewpoints is underlined. Next, rational planning approaches that "impose" goals, objectives and targets should be avoided and instead "means where intelligence and direction can emerge from ongoing organizational processes" should be fostered. In short Morgan sees organic structure, a bottom-up participatory approach and less formal ways of planning and reflection as core conditions for double-loop learning.

He completes the picture on organizational structures that enhance double-loop learning by referring to some concepts from systems theory.

The principle of *holographic systems* means that each part comprises a complete image of the whole. This metaphor emphasizes a certain redundancy in functions across subsystems and implies a more diverged authority systems; self-steering work teams can be seen as practical examples. Further following the metaphor of the organization as a human brain, strong interconnectivity between the subunits is emphasized. The principle of *requisite variety* places some boundary on the amount of redundancy (the degree to which units should be able to fulfil similar functions as others) in stating that "the internal diversity of any self-regulating system must match the variety and complexity of its environment". A practical implication is that organizations should pay close attention to the boundary relations between organizational units and their environments.

Apart from these two characteristics that bear on the structure of the organization there are two other principles that refer more to the procedural dimension of organizational functioning: the principles of *minimum critical specification* and *learning to learn.*

The principle of minimum critical specification bears some resemblance to the idea of subsidiarity, which popularly stated comes down to the principle that higher levels of an organizational structure should not do things that can also be carried out at a lower level. Similarly the principle of minimum critical specification speaks for limiting the pre-specification of organizational arrangements and processes to the maximum. In this way, "minimum critical specification suggests that managers and organizational designers should primarily adopt a facilitating and orchestrating role, creating 'enabling conditions' that allow a system to find its own form" (ibid., 101). Flexibility in organizational functioning is likely to result from such minimal management, which in its turn is seen as a favourable context for "inquiry driven action". The principle of *learning to learn* should prevent flexibility turning into chaos, and it is here that we are back with the organization's capacity for single- and double-loop learning.

Before dealing with the question about the relevance of the metaphor of the learning organization when applied to schools an attempt should be made to integrate the two conceptions of school organizations and school management, presented in the earlier sections on the professional bureaucracy and the rationally managed school, with educational leadership as a core construct. At first sight, the two perspectives provide considerable cognitive dissonance. How is the theoretically based image of the "professional bureaucracy", which also shows a lot of face-validity and common sense, to be reconciled with the empirically based concept of effective educational management?

In the first place, schools are nowadays not the exact copies of professional bureaucracies. Schools have been confronted with more demanding external requirements of both higher administrative levels and the consumers of education. In the "knowledge society", knowledge changes rapidly and there is a debate on whether to concentrate at teaching knowledge as such or rather strategies to acquire knowledge ("learning to learn"). As far as administration is concerned, in many countries, schools are given more autonomy in the domains of management and finance whereas—sometimes—there is less autonomy in the domain of the curriculum.

All these external changes work as pressures on the school to reconsider its functioning and perhaps even to change and innovate. And the importance of the role of the school head is now widely recognized. Another important reason why matters may start to depart from the picture of the professional bureaucracy is the availability of technology, Not only just teaching technology, like computer-assisted instruction, but also management and evaluation technology, in the form of school management information systems, pupil monitoring systems and school self-evaluation methods.

In the second place, "educational leadership" is not completely contrary to certain requirements of the professional bureaucracy. First, the educationally oriented school head can approach an individual teacher as a fellow-professional and colleague and, in this capacity discuss educational issues. Second, there can be a gradual implementation of meetings and work-sessions where teachers come together, and, in the presence of the head, discuss educational topics. The role of the

school head as an educational leader does definitely not preclude a democratic attitude and neither a collegial, supportive, coaching role. The point is that educational leadership by no means excludes a collegial, counselling-like approach, which would be more easily accepted by teachers. Third, the educational leader can opt for a management strategy that leaves the core of professional autonomy of teachers, namely, the process of teaching, largely as it is. This approach comes down to "freeing process and monitoring output" and can be seen as a form of functional decentralization at the school level.

Although in many respects the image of the school as a professional bureaucracy is still a valid image of the reality of school functioning it has moved in the direction of the rationally managed school, particularly because of an increased focus on results and outcomes and the standardization that follows the use of new technology.

Despite of the fact that not all of the features of the construct of the learning organization appear to be directly relevant and applicable to schools, the conclusion is that it is still a stimulating metaphor for effective school functioning, in a context that is partly standardized but also very much in movement. The core of the matter is organizational learning according to the cybernetic principle and negative feedback, which places monitoring and formative as well as summative evaluation at the centre of the idea of schools as learning organization.

Development of (in)effective School Culture: Self-organization and School Ecology

School effectiveness research has investigated the *school climate* in the sense of the general atmosphere at school. The term *school culture* expresses something similar. The meteorological metaphor is more often used to refer to the school environment as it confronts the students, whereas the anthropological metaphor of culture seems to be applied more generally to describe organizations. In this section, the term culture will be used, as it is closer to the central mechanism that is to be highlighted, namely, self-organization. When considering school climate, this is often seen as something that is planned and created, among others by the actions of the school leader, as well as something that emerges from the interactions among teachers and students, and students among themselves.

School culture can be defined as "the basic assumptions, norms and values, and cultural artefacts that are shared by school members, and which influence their functioning at school" (Maslowski 2001).

Schein (1985) distinguishes between three levels of organizational culture. The underlying level in Schein's classification consists of basic assumptions, which constitute in his view the essence of an organization's culture. This layer consists of basic assumptions shared by the organization's members, and which are more or less taken for granted: teachers are often no longer aware of the assumptions that

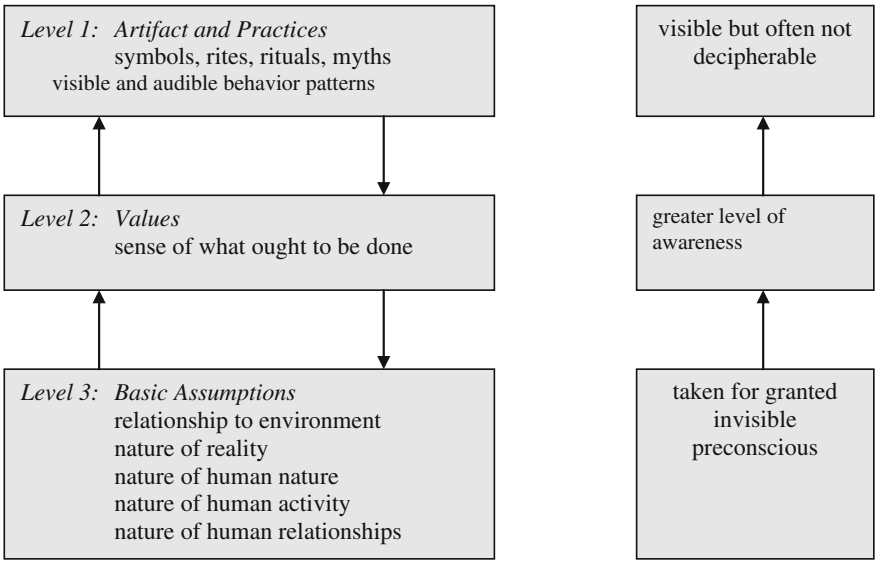

Fig. 4.3 Levels of culture and their interaction (adapted from Schein 1985, p. 14)

underlie their daily interpretations of their duties. These assumptions are likely to remain unconscious until another staff-member, student or parent challenges them. Schein distinguishes five groups of basic assumptions at the underlying level of culture: the organization's relationship to its environment, the nature of reality and truth, the nature of human nature, the nature of human activity and the nature of human relationships. For instance, the nature of human nature refers to whether humans are essentially "bad" or "good", and whether humans are basically fixed at birth, or whether they are "mutable and perfectible" (Schein 1985, p. 132; Maslowski 2001, p. 9).

At the second level of awareness, culture is defined in terms of values and norms. Values like collaboration or respect are often translated into norms for behaviour. Such behavioural norms, in fact, are "unwritten rules" according to which others are expected to behave. Norms also reflect what is considered to be *not done* in school, what is undesirable behaviour. Such norms may exist, for example, with regard to what teachers are expected to wear or not to wear, or what actions teachers are expected to take for their professional development (Maslowski 2001, p. 10) (Fig. 4.3).

The following long citation from Maslowski (2001, pp. 10–12) serves to clarify the most visible layer of organization culture, the one referring to *artefacts and practices*.

> "The third level in Schein's classification scheme consists of artifacts and practices. Ott (1989) has argued that both elements refer to essentially different components of an organization's culture. He therefore recommended distinguishing cultural artifacts, like symbols, heroes and myths, from behavioral patterns. Within these cultural artifacts, the

basic assumptions, values and behavioral norms of a school are 'visualized'. *Myths* articulate which past events have been important for members of the school. These 'critical' events are rendered in stories that are frequently recalled. Myths are often centered on actions or decisions taken by the *heroes* or *heroines* of the school. These people represent certain individual characteristics that reflect what members of the school value and serve as role models for the teachers. They may have been a founder of the school, the former principal, a charismatic teacher or even students who left school and whose actions exemplify the school's core values (Deal 1985; Van Hoewijk 1991). A third artifact pertains to *symbols* that exist in school. These indicate what meaning school members ascribe to various functions, parts or processes within school (Deal 1985).

Furthermore, in regard to behavioral aspects, the third level also consists of customs, rituals and procedures. In these practices or behaviors, the underlying assumptions, values and norms come to the surface. In each school certain behavioral patterns become established. These are not the result of any formal agreement or arrangement between teachers, but develop from socially accepted or reinforced behavior of the teachers (Deal 1985). *Customs* refer to 'the way we do things around here', which is often characteristic for the group of teachers within the school. Customs are culturally charged. Because certain 'ways of doing' exist, teachers and principals can predict in advance how others in school will react, what actions they will take and how they will perform their activities. Sometimes it is still possible to 'recognize' the beliefs or assumptions that led to the commonly accepted behavior in these customs. More often, however, these customs are so worn that they can only be interpreted in terms of shared assumptions, values and norms with great difficulty. Procedures, on the other hand, can often be interpreted more easily. To some extent, these procedures are prescribed by institutions outside the school, like the school district or the Ministry of Education. These procedures are less relevant from a cultural point of view. Much more relevant are the procedures that are developed *within* the school itself. These school-specific *procedures* reflect which actions have proven to be valuable for the school in the past and, therefore, have become institutionalized. From these procedures it can often be derived what is considered to be a 'good approach' in school. A final behavioral pattern that is culturally charged relates to school *rituals*. The term 'ritual' originates from the discipline of cultural anthropology, where it refers to the social customs around a certain event that has meaning for the members of a particular group. In schools, one may think of the ceremony that takes place when a teacher retires. These events take place according to a fixed protocol, consisting of several activities that may not be impressive from a substantial point of view, but which emphasize the solemnity of the event to the participants. Rituals, therefore, take place around events that are infused with meaning in the eyes of school members.

These three levels of culture are also referred to as the *cultural system* of a school. The cultural system, in turn, is made up of two components, latent and manifest cultural elements. *Latent* culture refers to tacit cultural levels. Generally the two inner levels, i.e. the underlying basic assumptions and the values and behavioral norms are referred to as latent elements of culture. The term *manifest* culture, on the other hand, refers to the visible part of the cultural system. More specifically, manifest culture refers to the myths, heroes and symbols of a school, and the established behavior patterns that have developed, like rituals, customs and procedures. Other characterizations for the manifest culture are 'cultural expressions' and 'cultural artifacts'.

To sum up, school culture is a generic term for the underlying assumptions, values and norms in school, and the myths, heroes, symbols, practices and rituals in which the latent culture manifests itself. The term 'cultural system', however, not only indicates that it covers a number of cultural elements, it also implies that the three levels are interrelated. This means that the basic assumptions of school members are related to their values and norms, and that these, in turn, are linked to the stories and symbols within the school and with the practices and rituals that exist".

Somewhat related to these levels of consciousness in organizational culture are Argyris and Schön's (1974) distinction between two types of action theories. A theory of action is defined as "a theory of deliberate human behavior which is for the agent a theory of control but which, when attributed to the agent, also serves to explain or predict his behavior" (ibid., p. 11). The two types of action theories they distinguish are "espoused theory" versus "theory-in-use". The espoused theory is the theory as communicated to others, the "official theory" one could say. The theory in use is more implicit and "might be inferred from organization members directly observable behavior" (ibid., p. 11).

The espoused theory is formulated at the level of the organization, in the sense of policy plans, designation of authority a task system and collectively agreed procedures. The theory of action is based on individual's behaviour in organizations and it may remain tacit. A reason for a theory-in-use remaining tacit could be that it is at odds with the espoused theory. Nevertheless, according to Argyris and Schön, "the largely tacit theory-in-use accounts for organizational identity and continuity".

Both concepts, Schein's conceptualization of the organization's culture and Argyris and Schön's distinction of two kinds of organizational action theories have in common that they distinguish a more explicit formal level of consciousness of regularities in the organization's functioning next to a more implicit, semi-conscious, tacit images of organizational functioning.

This last level is considered as making up the more fundamental identity of the organization, which goes back to beliefs, attitudes and cognitive styles of the individual members. A methodological implication seems to be that reconstruction of this deeper level of organizational functioning largely appears as an aggregation of the attitudes, beliefs and behavioural maps of individual members of the organization.

Given the large number of elements in the sense of individual members (subsystems) and the variety of norms, methods and goals (aspect systems) as well as temporal dynamics (phase-systems) and thus the enormous variety of possible interactions, the "coming to be" of an organization's culture should be seen as something very complex.

The link between endogenous development "autopoiesis" and this semi-conscious level of organizational functioning and cultural identity is to be found in this very complexity, as will be pointed out after further explication of these self-referential processes.

The "implicate order" at this level is seen by the physicist Bohn as a creative process, which like a hologram, has everything enfolded in everything else (Bohn, cited by Morgan 1986, p. 234). The regenerative processes in this "implicate order" are seen as "logics of change" that help to explain the concrete form of the world at any given point in time (ibid., p. 234).

In analogy to the discussion on classroom ecology and classroom climate, one could view the organization culture as closely connected to composition effects, in this case the composition of personality traits of teachers in a school. With reference to the development of an evaluation culture at school, Scheerens (2004) mentions

Personality trait	Related behavioral disposition
Anxiety	Evaluation apprehension; ambiguity tolerance
Egoism	Opportunism
Introversion	Degree of openness
Locus of control	Sense of efficacy
General Intelligence	Information processing capacity

Fig. 4.4 Personality traits and dispositions relevant to teachers' attitudes towards evaluation and assessment, from Scheerens (2004)

the following personality traits of teachers as relevant entrance conditions (Fig. 4.4).

When considering some cultural orientations inherent in the core factors that are identified by school effectiveness research, the factors "high expectations" and "achievement orientation", could be seen as partly determined by staff composition on locus of control and experienced sense of efficacy.

Seen in this way, the development of a school culture is less of a managed process, but more like a process of endogenous development, whereby the shaping of the latent culture of organizations has two main elements:

(a) the assumption that aggregates of relevant organization member variables shape the 'sense' or the initial frame of reference of the latent organization culture;
(b) the assumption that this frame of reference sets in motion circular, self-referential processes that reinforce the initial pattern.

The operational implication for research would be to try and measure relevant personality background characteristics of teachers, and study their impact both at teacher level as an antecedent condition to classroom teaching and at school level, as a composition effect and a determinant of school culture and climate.

The school culture can also be analysed as a micro-cosmos for informal learning at school, where students are confronted with normative aspects of how the school is run, the way different functionaries interact and communication is regulated. Such exemplary behaviour could be studied as a relevant context for the development of active citizenship.

Schools as High-Reliability Organizations: Inspirations from Neo Institutional Economics

The development of the latent school culture as a process of internal endogenous development emphasizes the Praxis dimension of Fend's scheme that was discussed at the beginning of the section on theories of the school, particularly on the leeway for autonomy at school level. The exact opposite is the perspective on the functioning of schools as depending on institutional norms and formal regulations. These views are fueled by neo institutional theory emphasizing juridical norms and fixed definitions on the way things are to be done. In terms of organizational models, this emphasis on formal regulation has correspondence to the classical theory of the bureaucracy. Perfection of formal procedure is also inherent in quality management schemes as the well-known ISO norms. McMeekin (2003) takes neo institutionalism as a source of inspiration to define the "institutional climate" at school as the key feature of effective schooling. He states that "the institutional climate of schools, which includes formal rules, informal rules, mechanisms for enforcing both kinds of rules, clear objectives and an atmosphere of cooperation and trust, has a strong influence on school performance".

A fascinating piece of conceptual work and related empirical investigation is provided in Stringfield's description of "high reliability organizations" (Stringfield 1995; Stringfield et al. 1995).

The principles of high-reliability organizations (e.g. nuclear power plants and air navigation systems) are:

1. the notion that failures within the organization would be disastrous;
2. clarity regarding goals and a strong sense of the organization's primary mission held by the staff;
3. use of standard operating procedures (e.g. scripts);
4. importance of recruitment and intensive training;
5. initiatives that identify flaws (e.g. monitoring systems);
6. considerable attention to performance, evaluation and analysis to improve the processes of the organization;
7. monitoring is sees as mutual, without counterproductive loss of overall autonomy and confidence;
8. alertness to surprises or lapses (notion that small failures could cascade into major system failures);
9. hierarchical structure, allowing for collegial decision making during times of peak loads;
10. equipment is maintained in the highest working order;
11. high-reliability organizations are invariably valued by their supervising organizations;
12. "short-term efficiency takes a back seat to high reliability" (from Stringfield 1995, pp. 83–91).

In both the evaluation of major effectiveness-oriented improvement projects in the USA and the evaluation of a highly structured primary school program (the Calvert-Barclay school project) in the UK, evidence was found that supported the validity of the high-reliability organization's image. The Calvert-Barclay project is particularly illustrative. It describes the implementation of a highly structured and traditional academically oriented private school program in an inner-city school. The success of the program in these two strongly divergent settings provides additional support to the generalizability of this structured approach.

Despite well-known criticisms of the usefulness of rational planning and mechanistic structuring approaches in educational organizations (e.g. Lotto and Clark 1986), these latter examples show that a plea can be made for formalized and highly structured educational programs, supported by structures that emphasize order, coordination and unity of purpose. The major challenge seems to be to combine effectively standardized procedures and partial mechanistic structuring to conditions that nevertheless are sufficiently motivating to educational professionals and likewise keep appealing to the creative insights of all the members of the organization.

The formalization dimension is closely related to rules and inputs that are specified at above school levels; and in that realm the curriculum is particularly important. Some research results appear to suggest that centralization in the domains of curriculum and assessment is more effective than decentralization. Willms and Somers (2001) point at the superior results of Cuba in an international comparative assessment among nine Latin American countries; Bishop (1997) shows correspondence between countries having a standard-based examination at secondary level and student performance, while Woessmann (2000, 2004) and Fuchs and Woessmann (2004) have presented results that point at the superiority of centralized curricula combined with school autonomy in other managerial domains.

At school level, the degree of leeway that teachers have vis à vis directive leadership on the one hand, and formally prescribed programs on the other is a contested area as well. Concepts like "teacher leadership" and "empowerment" are an issue in the literature on educational reform, but can also be studied as variables in the day to day running of schools and compared for their effectiveness. "Expanded teacher leadership roles range from assisting with the management of schools to evaluating educational initiatives and facilitating professional learning communities", say York-Barr and Duke (2004, p. 255), and they go on to say that "The hope for teacher leadership is continuous to improvement of teaching and learning in our nation's schools, with the result being increased achievement with every student".

The opposite position is manifested in Comprehensive School Reform (CSR) projects, where "externally" developed CSR models provide a type of top-down direction for designing and supporting the process of school reform (Borman et al. 2003, p. 126). Where the movement for teacher leadership expresses hope for bottom-up reform, the CSR approach also shows rather impressive claims. To provide another citation of Borman et al.: "The problem is that the complex educational changes demanded by current standards-based reform initiatives,

combined with an increasingly heterogeneous student population largely composed of students whom schools traditionally have failed, have pushed the technology of schooling to unprecedented levels of complexity. In many ways, expecting local educators to reinvent the process of educational reform, school by school, is both unrealistic and unfair" (ibid., 126).

The empirical evidence on the results of CSR programs points the attention at pre-structuring and regulation of teaching by explicit, externally developed guidelines, and in this way show a more circumscribed aspiration of the teaching profession than the followers of the "empowerment paradigm" (Muijs and Reynolds 2001) would have it.

Concluding Comments on School Theories and Models: What Are the Most Important Factors and How Would They Interact?

Each of the five school models emphasizes certain variables; this is summarized in Table 4.3.

Table 4.3 School models and key variables

School model	Key variables
Professional bureaucracy	Initial teacher training, professional development
Effective schools model	Achievement, orientation, high expectations
	Educational leadership
	Consensus and cooperation of staff
	Curriculum quality and opportunity to learn
	School climate
	Evaluative potential
	Parental involvement
	Effective learning time
School as learning organizations	Professional development
	Evaluation and feedback
School culture as a product of self-organization	Characteristics of teachers
	Student composition effects
	Teacher composition effects
	Complex interactions among teachers
Schools as high reliability organizations	Externally developed curricula and lesson plans
	Frequent monitoring
	Institutional climate of schools

The summary table indicates that most key variables that are highlighted in the four other models are included in the effective schools model as well. Another way of putting this would be to say that these four other models, the professional bureaucracy, the learning organization the culture/self-organization model and the model of the school as a high-reliability organization provide a theoretical basis for specific variables that have been incorporated in effective schools research.

A final question could be about the way these key factors of effective schooling are expected to interact. A partial answer to this was given in the presentation of the indirect causation of educational leadership through a set of hypothetical intermediary variables. Depending on one's theoretical point of departure, different configurations might be proposed as well, however. From the perspective of self-organization one might perhaps see the leadership style more as an effect than a "cause" of organizational culture, which, in its turn, might be strongly influenced by the staff composition and pre-dominant attitudes resulting from that.

Next, one could look for certain interrelationships and partial overlap between some variables. Achievement orientation, high expectations and a positive climate share a similar orientation that could be both planned and emergent from individual and compositional background characteristics of teachers and students. Opportunity to learn and achievement orientation (towards basic subjects) could be seen as manifestations of *well-targeted* schooling, whereas effective learning time would be associated with both instructional leadership and disciplinary climate, and perhaps also with coordination and consensus, as a mechanism that optimizes connectivity between grades and classes. With respect to high expectations, one should also be aware reciprocal causation, in the sense that high expectations could be seen as both the cause and the effect of high performance.

Bearing in mind these interrelationships, the set of effectiveness enhancing conditions could also be taken all together and studied or optimized as a total set, as is the case in CSR Projects. CSR is defined on the basis of 11 components by the US Department of Education, cf. Borman et al. (2003, p. 127):

1. "Employs proven methods of student learning, teaching, and school management that founded on scientifically based research and effective practices and have been replicated successfully in schools;
2. Integrates instruction, assessment, classroom management, professional development, parental involvement, and school management:
3. Provides high-quality and continuous teacher and professional development and training;
4. Includes measurable goals for student academic achievement and establishes benchmarks for meeting those goals;
5. Is supported by teachers, principals, administrators, and other staff throughout the school;

6. Provides support for teachers, principals, administrators, and other school staff, by creating shared leadership and a broad base of responsibility for reform efforts;
7. Provides for meaningful involvement of parents and the local community in planning, implementing and evaluating school improvement activities;
8. Uses high-quality external technical support and assistance from an entity that has experience and expertise in school wide reform and improvement, which may include and institution for higher education;
9. Includes a plan for the annual evaluation of the implementation of the school reforms and the student results achieved;
10. Identifies the available federal, state, local, and private financial and other resources that schools can use to coordinate services that support and sustain the school reform effort; and
11. Meets one of the following requirements: Either the program has been found, through scientifically based research, to significantly improve the academic achievement of participating students; or strong evidence has shown that the program will significantly improve the academic achievement of participating students."

A final source of evidence for the coherence and synthetic nature of the set of effectiveness enhancing conditions highlighted throughout this section is the finding that highly effective schools seem to do well on all or most of these factors, while failing schools do badly on all or most of them (Stringfield 1998).

Conclusion About the Effective Functioning of Schools

School leadership, school policies and created organizational conditions at school level are seen as malleable conditions that are expected to facilitate good quality teaching conditions and, in this way to stimulate student learning indirectly (mediated by teaching, and a productive classroom climate) and directly by means of a generally positive and well-regulated learning climate for each student. These malleable school conditions are seen as shaped, stimulated and constraints by school antecedents that are externally determined, and by school ecological factors that are interactions of internally and externally determined aspects. The associations between the three main components (school antecedents, school ecology and active school policies, management and organization) are depicted in Fig. 4.5.

Finally, the key variables are once more resumed in Summary Table 4.4.

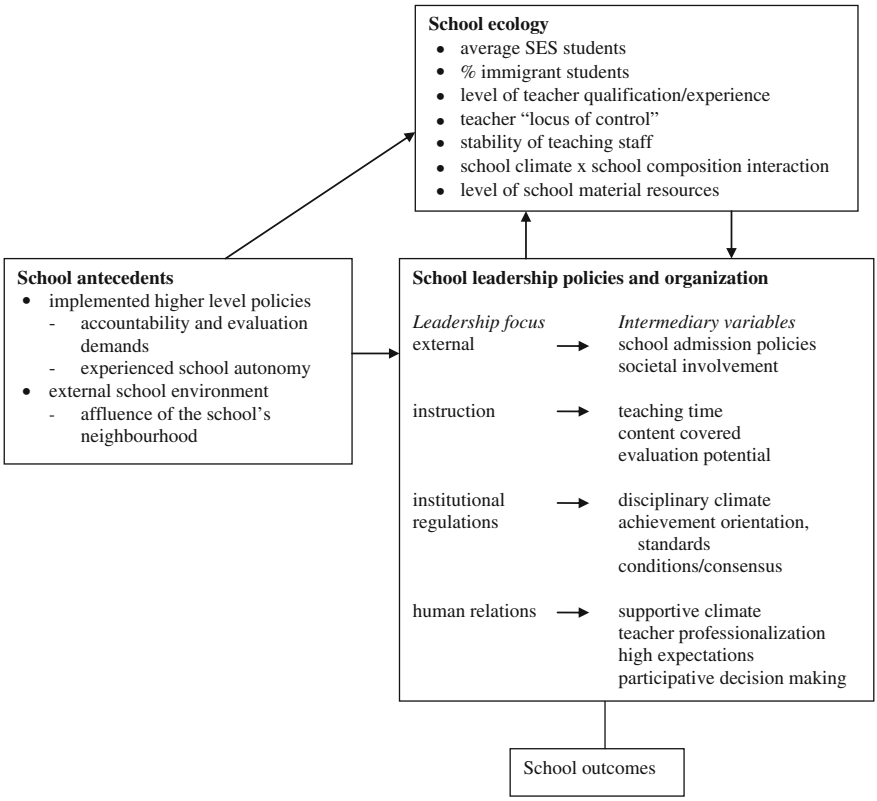

Fig. 4.5 School functioning

Table 4.4 School variables

School antecedents	School ecology	School leadership, policies and organization
Implemented higher level policies	– Student composition (e.g. school average SES)	– Leadership focus; specifically the degree of instruction- oriented leadership;
• Accountability and evaluation demands	– Percentage of students from immigrant background;	– Achievement orientation/high expectations
• Experienced school autonomy	– Percentage of students with a special education profile	– Teaching time
External school environment	– Teacher composition (e.g. the average qualification, experience and locus of control scores of the teachers)	– Quality of school curriculum, opportunity to learn;

(continued)

Table 4.4 (continued)

School antecedents	School ecology	School leadership, policies and organization
• Affluence of the school neighbourhood	– Stability of teaching staff over the last 5 years	– Coordination, cooperation and consensus among staff;
	– School climate x school composition interaction	– A safe and orderly climate, supporting positively experienced interactions
	– Level of school material resourcing	– Parental involvement
	– Level of extra-curricular activities of the school	– Opportunities for professional development of teachers
		– School admission policies
		– Evaluation, feedback and monitoring
		– Degree to which teachers participate in decision making

References

Argyris, C., & Schön, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.

Bishop, J. (1997). The effect of national standards and curriculum-based exams on achievement. *The American Economic Review, 87*(2), 260–264.

Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230.

Cotton, K. (1995). *Effective schooling practices: A research synthesis. Update*., School improvement research series Portland: Northwest Regional Educational Laboratory.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London, New York: Routledge.

Deal, T. E. (1985). The symbolism of effective schools. *Elementary School Journal, 85*(5), 601–620.

Elmore, R. F. (2000). *Building a new structure for school leadership.* Washington DC: The Albert Shanker Institute.

Fend, H. (1981). *Theorie der Schule*. München: Urban & Schwarzenberg.

Fend, H. (2006). *Neue Theorie der Schule*. VS Verlag für Sozialwissenschaften.

Fuchs, Th., & Woessmann, L. (2004). *What accounts for international differences in student performance? A re-examination using PISA data*. CESifo Working paper No. 1235. Munich: University of Munich.

Hallinger, P., & Heck, R. H. (2010). Leadership for learning: Does collaborative leadership make a difference? *Educational Management Administration & Leadership, 38*(6), 654–678.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Hoewijk, P. M. Th. van (1991). Schoolcultuur. In P. J. J. Stijnen & J. F. M. Claessen (Eds.), *Leerboek Schoolmanagement* (pp. 83–102). Heerlen: Open Universiteit.

Huber, S. G., & Muijs, D. (2010). School leadership effectiveness. *Studies in Educational Leadership*, *10*, 57–77 (Springer).

Irwin, C. C. (1986). What research tells the principal about educational leadership. *Scientica Paedagogica Experimentalis, 23*, 124–137.

Kerr, St., & Jermier, J. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behaviour and Human Performance, 22*, 375–403.

Lotto, L. S., & Clark, D. L. (1986). Understanding planning in educational organizations. *Planning and Changing, 19*, 9–18.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Leithwood, K. A., & Montgomery, D. J. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research, 52*, 309–399.

Levine, D. K., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, Wise: Nat. Centre for Effective Schools Research and Development.

Luyten, H. (2009). School leadership and teachers' self-efficacy: The development and testing of a causal model. In: G. ten Bruggencate, H. Luyten, & J. Scheerens (Eds.) Quantitative analyses of international data, exploring indirect effect models of school leadership. Report for the EU-funded LISA (Leadership and International Student Achievement). Enschede: University of Twente.

Maslowski, R. (2001). *School culture and school performance*. An explorative study into the organizational culture of secondary schools and their effects. dissertation, Enschede: Twente University Press.

McMeekin, R. W. (2003). *Incentives to improve education: A new perspective*. Cheltenham: Edward Elgar.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Morgan, G. (1986). *Images of organizations*. Beverly Hills: Sage.

Muijs, D., & Reynolds, D. (2001). *Effective teaching*. London: Paul Chapman.

Ott, J. S. (1989). *The organizational culture perspective*. Pacific Grove, CA: Brooks/Cole.

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., & Teddlie, C., et al. (2014). Educational Effectiveness Research (EER): A state of the art review. *School Effectiveness and School Improvement*, *25*, 197–230 (2012).

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J. (2004). The evaluation culture. *Studies in Educational Evaluation, 30*, 105–124.

Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.

Scheerens, J., Glas, C., & Thomas, S. (2003). *Educational evaluation, assessment and monitoring*. Lisse: Swets & Zeitlinger.

Scheerens, J. (Ed.). (2012). *School leadership effects revisited. Review and meta-analysis of empirical studies.* Heidelberg: Springer.

Scheerens, J. (2014). School, teaching, and system effectiveness: Some comments on three state-of-the-art reviews. *School Effectiveness and School Improvement, 25*(2), 282–290.

Schein, E. H. (1985). *Organizational culture and leadership: A dynamic view*. San Francisco, CA: Jossey-Bass.

Stringfield, S. C., & Slavin R. (1992). Raising societal demands, high reliabilty organizations, school effectiveness, "Succes for All", and a set of modest proposals. Enschede: Address to the ICO-symposium.

Stringfield, S. (1995). Attempting to enhance students' learning through innovative programs: The case for schools evolving into high reliability organizations. *School Effectiveness and School Improvement, 6*(1), 67–96.

Stringfield, S. (1998). An anatomy of ineffectiveness. In L. Stoll & K. Myers (Eds.), *No quick fixes: Perspectives on schools in difficulties* (pp. 209–221). London: Falmer.

Stringfield, S. C., Bedinger, S. & Herman R. (1995). *Implementing a private school program in an inner-city public school: Processes, effects, and implications from a four year evaluation*. Paper presented at the ICSEI Conference, Leeuwarden, The Netherlands.

ten Bruggencate, G., Luyten, H., Scheerens, J., & Sleegers, P. (2013). Modeling the influence of school leaders on student achievement: How can school leaders make a difference? *Educational Administration Quarterly, 48*(8), 699–732.

Van der Werf, M. P. C. (1988). *Schoolwerkplanontwikkeling in het basisonderwijs* (*School development plans in primary education*). Groningen: University of Groningen.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*, 1–19.

Weick, K. E. (2001). *Making sense of organizations*. Chichester: Wiley.

Willms, J. D., & Somers, M. A. (2001). Family, classroom, and school effects on children's educational outcomes in Latin America. *School Effectiveness and School Improvement, 12*(4), 409–445.

Woessmann, L. (2000). *Schooling resources, educational institutions, and student performance: The international evidence*. (Kiel Working paper No. 983). Kiel, Germany: Kiel Institute of World Economics.

Woessmann, L. (2004). *The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA*. CESIFO Working paper no. 1330, University of Munich, Munich, Germany.

York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? From two decades of scholarship. *Review of Educational Research, 74*(3), 255–316.

Chapter 5
School Effectiveness Research and the Meaning of the Most Important Effectiveness-Enhancing Conditions

Keywords School effectiveness · Achievement orientation/high expectations/ teacher expectations · Educational leadership · Consensus and cohesion of staff · Curriculum quality and opportunity to learn · School climate · Evaluative potential · Parental involvement · Classroom climate · Effective learning time (classroom management) · Structured instruction · Differentiation/adaptive instruction · Feedback and reinforcement

Introduction: The Overall Design of School Effectiveness Studies

Similar to the description of educational effectiveness in previous chapters, the elementary design of school effectiveness research is the association of hypothetical effectiveness-enhancing conditions of schooling and output measures, mostly student achievement. The basic model from systems theory that was introduced in Chap. 1 is helpful to clarify this basic design. The major task of school effectiveness research is to reveal the impact of relevant input characteristics on output and to "break open" the black box in order to show which process or throughput factors "work", next to the impact of contextual conditions. Within the school it is helpful to distinguish a school and a classroom level and, accordingly, school organizational and instructional processes. To facilitate the use of this chapter as a stand-alone paper, the basic systems model from Chap. 1, is reproduced once more in Fig. 5.1.

In Fig. 5.2 is a somewhat different presentation of the systems model is presented. Here the central box is defined at the level of an organization, in our case, a school. The functioning of the organization consists of inputs flowing into the central box and by outputs being "somehow" produced (see Fig. 3.2).

In Fig. 5.2 it is assumed that within the black box *processes* take place that transform inputs into outputs. When it is attempted to further describe these processes in terms of which process characteristics are most effective in obtaining

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_5

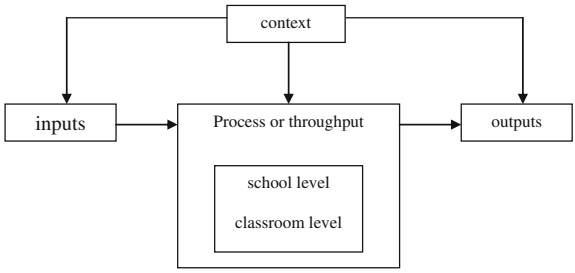

Fig. 5.1 A basic systems model on the functioning of education

input → organisation as a black box → output

Fig. 5.2 The organization as a *black box*

desired levels of outputs, the model of Fig. 5.1 gets filled with content. In this way the model is often used as a conceptual framework to summarize the results of school and instructional effectiveness research and examples of these have been shown in Chaps. 2, 3 and 4. An earlier example in which this approach was followed is the model by Scheerens (1989), shown in Fig. 5.3.

The notion of quality inherent in integrated school effectiveness models like the one in Fig. 5.3 is that:

(a) outputs are the basic criteria to judge educational quality;
(b) in order to be able to properly evaluate output, achievement or attainment measures should be adjusted for prior achievement and other pupil intake characteristics; in this way the value added by schooling can be assessed;
(c) in selecting variables and indicators to assess processes and context one should look for those factors that have been shown to be correlated with relatively high output, adjusted in terms of "added value" as described above;
(d) the model is a multi-level model, uniting effectiveness-enhancing conditions at system, school, classroom and individual student level.

It should be noted that educational effectiveness models do not prescribe the types of outputs that should be used to assess quality. In principle all types of outputs, cognitive or non-cognitive could be inserted in the right-hand box of Fig. 3.3. In the actual practice of school effectiveness research, however, cognitive outcomes, mostly in terms of achievement in core subjects like reading, arithmetic and language, have predominated. The process factors shown in the middle section of Fig. 5.3 might be somewhat different if non-cognitive outcomes or less subject matter tied cognitive outcomes would have been used in the actual research studies.

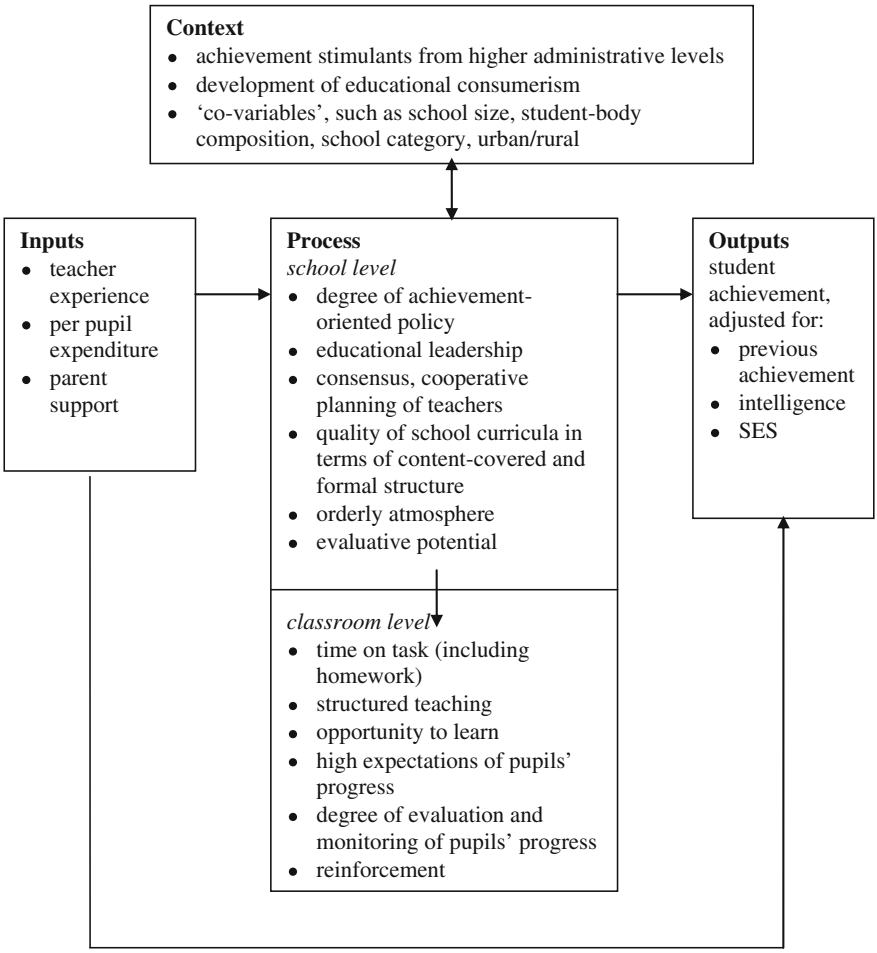

Fig. 5.3 A summary of the findings from school effectiveness research, from Scheerens (1989)

Different Strands of Educational Effectiveness Research

Research tradition in educational effectiveness varies according to the emphasis that is put on the various antecedent conditions of educational outputs. These traditions also have a disciplinary basis. The common denominator of the five areas of effectiveness research that will be distinguished is that in each case the elementary design of associating outputs or outcomes of schooling with antecedent conditions (inputs, processes or contextual) applies. The following research areas or research traditions will be considered in summarizing the research results obtained in developed countries:

Table 5.1 General characteristics of types of school effectiveness research, cited from Scheerens and Bosker (1997, p. 140)

	Independent variable type	Dependent variable type	Discipline	Main study type
a. (Un)equal opportunities	Socio-economic status and IQ of pupil, material school characteristics	Attainment	Sociology	Survey
b. Production functions	Material school characteristics	Achievement level	Economics	Survey
c. Evaluation compensatory programmes	Specific curricula	Achievement level	Interdisciplinary pedagogy	Quasi-experiment
d. Effective schools	"Process" characteristics of schools	Achievement level	Interdisciplinary pedagogy	Case study
e. Effective instruction	Characteristics of teachers, instruction, class organization	Achievement level	Educational psychology	Experiment observation

(1) Research on equality of opportunities in education and the significance of the school in this.
(2) Economic studies on education production functions.
(3) The evaluation of compensatory programmes.
(4) Studies of unusually effective schools.
(5) Studies on the effectiveness of teachers, classes and instructional procedures.

In developing countries there is a strong predominance of studies of the education production function type. Relatively few of these have been expanded by including school organizational and instructional variables. Table 5.1 summarizes the main characteristics of the five research traditions.

In the historical development of school effectiveness studies these various approaches to educational effectiveness have become integrated. Integration was manifested in the conceptual modelling and the choice of variables. At the technical level multi-level analysis has contributed significantly to this development. In contributions to the conceptual modelling of school effectiveness, schools became depicted as a set of "nested layers" (Purkey and Smith 1983), where the central assumption was that higher organizational levels facilitated effectiveness-enhancing conditions at lower levels (Scheerens and Creemers 1989). In this way a synthesis between production functions, instructional effectiveness and school effectiveness became possible. This was achieved by including the key variables from each tradition, each at the appropriate "layer" or level of school functioning [the school environment, the level of school organization and management, the classroom level and the level of the individual student]. Conceptual models that were developed according to this integrative perspective are those by Scheerens (1989), Creemers (1994) and Stringfield and Slavin (1992).

Results of Integrated Multi-level School Effectiveness Studies

There is a fairly large consensus on the main categories of variables that are distinguished as effectiveness-enhancing conditions in the reviews, also when earlier and more recent reviews are compared. Scheerens and Bosker (1997, 156) summarize the characteristics listed in the reviews by Purkey and Smith (1983), Scheerens (1992), Levine and Lezotte (1990), Sammons et al. (1995), Cotton (1995).

Consensus is largest with respect to the factors:

- achievement orientation (which is closely related to "high expectations");
- cooperation;
- educational leadership;
- frequent monitoring;
- time, opportunity to learn and "structure" as the main instructional conditions.

This consensus is maintained in more recent research review, most notably the one by Reynolds et al. summarized in Table 4.1 of Chap. 4.

Behind this consensus on general characteristics hides considerable divergence in the actual operationalization of each of the conditions. Evidently, concepts like "productive, achievement-oriented climate" and "educational leadership" are complex concepts and individual studies may vary in the focus that different elements receive.

Scheerens and Bosker (1997, Chap. 4) provide an analysis of the meaning of the factors that are considered to work in schooling apparent from the actual questionnaires and scales as used in ten empirical school effectiveness studies.

The Meaning of the Key Effectiveness-Enhancing Conditions

Table 5.2 lists the variables that have received relatively most support in empirical school effectiveness research. It should be noted that the overview contains four factors that are basically defined, not at the school, but rather at the classroom level. In some applications, however, they may also have a school-level interpretation, in the sense of aggregated classroom information, and school policies with preference for a certain teaching approach, like structured teaching or adaptive teaching.

The elements found in the operational definitions and instruments concerning these factors will be summarized for each factor, but first an impressionistic view on the conceptual "core" of each factor will be given.

Achievement Orientation/High Expectations

Within the set of operational definitions that was considered the following main components could be distinguished:

Table 5.2 General effectiveness-enhancing factors

1.	Achievement orientation/high expectations/teacher expectations
2.	Educational leadership
3.	Consensus and cohesion among staff
4.	Curriculum quality/opportunity to learn
5.	School climate
6.	Evaluative potential
7.	Parental involvement
8.	Classroom climate
9.	Effective learning time (classroom management)
10.	Structured instruction
11.	Differentiation, adaptive instruction
12.	Feedback and reinforcement

- A clear focus on the mastery of basic subjects.
- Fostering high expectations on pupils' achievement, at school and teacher level.
- The use of records on pupils' progress.

Other elements of pressure for achievement that are not contained in this overview, but have been mentioned in the literature are:

- "placing 'attainment' on the agenda of staff meetings and in talks between the school head and individual staff";
- "employing achievement pressure as a criterion when recruiting new teaching staff";
- "implementing resources, including testing systems, that make it easier to introduce an achievement-oriented policy" (Scheerens 1992, p. 87).

The general concept of achievement orientation and fostering high expectations comprises overt policy choices, attitudes, behaviours and structural facilities. The core idea is the determination to get from pupils what they are worth, in terms of aptitudes and home environment. Standard setting in a way that pupils are challenged, but not demotivated because the standards are either too high or too low appears to be the main structural measure in a "balanced" interpretation of achievement orientation. "Balanced" in the sense that no monomaniacal preoccupation with achievement, regardless of ability levels, is implied, but care is taken of individual differences between pupils.

Educational Leadership

An extensive overview of the conceptual development on school leadership has been provided in the previous chapter. Therefore only two summary schemes, cited from Scheerens 2012, and from Hendriks and Scheerens (2013) are presented below.

A schematic overview of the development in the concept formation on school leadership is presented in Table 5.3.

In Table 5.4 core leadership functions are matched with more specific leadership behaviour.

Table 5.3 Concepts of leadership at school

Type of leadership	Orientation
Instructional leadership	Curriculum and instruction
Extended instructional leadership	School mission
	Managing the curriculum
	Providing learning climate
Transformational leadership	Models organizational values
	Develops shared mission
	Provides intellectual stimulation
	Builds consensus
	Redesigns organizational structure
Integrated leadership	Conditions supporting school improvement
	Instructional leadership; broader perspectives on organizational effectiveness; leadership roles "delegated" to people and structural coordination mechanisms

Table 5.4 Leadership functions and corresponding leadership behaviours

Leadership functions	Leadership behaviour
Developing a vision	External contacts
	Buffering
	Setting values
Managing the teaching and learning programme	Direction setting (vision, goals, standards)
	Monitors curriculum and instruction (managing the instructional programme)
	Redesigning the organization
Understanding and developing people	HRM and HRD
	Coaches teachers
	Recruits teachers
	Builds consensus
	Individual support
	Intellectual stimulation
Redesigning the organization	Uses "substitutes" for leadership, Distributes leadership tasks
	Creates climate

Consensus and Cohesion Among Staff

Given the traditional autonomy of teachers it is clear that consensus, cohesion and sufficient continuity for pupils when they pass from one teacher to the next, should not be taken for granted in schools. Therefore, in many school effectiveness studies, the degree to which schools succeed in building coherence and consistency is seen as a hypothetical explanation for the fact that some schools do better than others.

In the operational definitions and instruments that were analysed the following components of consensus and cooperation were distinguished:

- Types and frequency of meetings and consultations.
- The contents of cooperation.
- Satisfaction about cooperation.
- The importance attributed to cooperation.
- Other indicators of successful cooperation.

In the way consensus and cooperation is measured, facts, actual cooperation and frequency of sessions where staff meet and cooperate, as well as perceptions and attitudes on cooperation are all included. With respect to the substance of cooperation both agreement on overall mission and educational philosophy as well as consultation on "technical" aspects of teaching and instruction are measured.

There appears to be no agreement on areas of cooperation that are thought to be particularly relevant. Across studies a broad range of cooperation activities and "cooperation content" are chosen.

Curriculum Quality and Opportunity to Learn

The curriculum has been described as the "blueprint" for the functioning of the primary process in education. In articulating the curriculum and by indicating clear targets, the curriculum could function as a powerful coordination mechanism (i.e. a form of standardization). On the other hand such standardization is usually balanced by opportunity for teachers to exercise their own professional autonomy.

The degree to which content that is actually taught (sometimes described as the "implemented curriculum") corresponds to the test or examination of items used to assess achievement (the achievement curriculum) is usually taken into account in international comparative studies under the label "opportunity to learn".

Examination of the instruments in this area led to the following categories:

- The way curricular priorities are set.
- Choice of methods and textbooks.
- Application of methods and textbooks.
- Opportunity to learn.
- Satisfaction with the curriculum.

Further analysis of the literature reveals that the core elements appear to be:

- a clear focus of the curriculum;
- coordination and alignment of the curriculum (relationship goals and curricular choices, correspondence among grade levels, classes and teachers);
- test curriculum overlap, or "opportunity to learn".

School Climate

In the history of school effectiveness research two aspects of school climate have received emphasis: orderliness and achievement orientation. In the earlier presentation achievement orientation was treated as a characteristic of explicit or even official policy. Achievement-oriented climate refers more to internalized norms and views of individual staff members shared with their colleagues, also in less formal relationships. A third aspect of school climate is the experience of the general "goodness" of all kinds of internal relationships and the satisfaction this give to staff and pupils.

Indicators on the school climate range from perceptions and normative views to behavioural characteristics and factual circumstances like a set of explicit behavioural rules, absenteeism statistics and characteristics of the school building.

Rules about proper behaviour and discipline express the conviction and effort of schools to suppress disruptive and negative, non-task related activities as much as possible. In school effectiveness thinking "good relationships" and satisfaction are considered instrumental to enhanced school effectiveness, and not just as "aims in themselves".

The main sub-categories express the breadth of scope of the school climate concept.

Evaluative Potential

The concept of "evaluation potential" (Scheerens 1987) expresses the aspirations and possibilities of schools to use evaluation as a basis for learning and feedback at the various levels within the organization, also taking into account limitations and constraints.

Aspects of this concept are:

- priority given to assessment and monitoring;
- evaluation technology (e.g. standardized pupil monitoring systems or computerized "test service systems");
- use of evaluation results and records at the school level.

One of the problems in measuring schools' involvement in evaluation is the diversity in evaluation methods, which range from very informal procedures like the marking of assignments to the regular use of standardized achievement tests. Also, there are several objectives of school-based evaluation:

- monitoring of "normal" progress in pupils' achievement;
- diagnosing learning difficulties;
- assessment of whole school, department or classroom/teacher performance;
- school diagnosis as a basis for prospective innovations and school improvement activities;
- assessment to meet external accountability requirements;
- assessment to be used as a basis for "marketing" the school and informing parents and other stakeholders.

The main aspects of "evaluative potential", distinguished in the introductory section on this factor, orientation, technique and use, are clearly reproduced in the instruments that were analysed.

Parental Involvement

Continuity in home and school learning and an active involvement of parents in school matters is considered relevant in various strands of school effectiveness research. Both actual involvement and effort of the school to facilitate involvement are usually included in instruments for measuring this alleged effectiveness-enhancing factor.

Classroom Climate

Like in school climate orderliness, good relationships and satisfaction are the main components of classroom climate.

In comparison to the components that were distinguished in "school climate" the achievement orientation component is missing. This aspect, however, is more or less covered in another factor, namely teacher expectations.

Effective Learning Time

Learning time can be interpreted as a measure of the quantity of exposure to "educational treatment" at school. Time can be assessed at school and at classroom level, and a distinction is to be made between "planned time" (e.g. the time per subject matter area in the timetable) and "implemented time" or "time on task".

When summarizing the elements found in the set of instruments that were analysed the following components were distinguished:

- Importance of effective learning time.
- Monitoring of absenteeism.
- Time at school level.
- Time at classroom level.
- Classroom management (avoiding and minimizing ineffective "time consumers").
- Homework.

Structured Instruction

Although there are diverging instruction–theoretical and pedagogical perspectives on "good instruction", (see Chaps. 1 and 2) in school effectiveness research the view that instruction should be well-structured and closely monitored predominates. The following components could be distinguished:

- Importance of structured instruction.
- Structure of lessons.
- Preparation of lessons.
- Direct instruction.

The main sub-factors in "structured instruction" are basic requirements of well-prepared and well-controlled teaching on the one hand and aspects of direct instruction on the other.

Differentiation

Differentiation is aimed at instruction that is adaptive to the specific needs of subgroups of pupils. The success of differentiation is to a large extent dependent on school and classroom organization. Crucial intervening variables are time on task, and the quality of tuition during group work.

Reinforcement and Feedback

Reinforcement and feedback are important basic conditions for learning. It should be noted that reinforcement and feedback have both cognitive and motivational implications, as a basic requirement in learning and in rewarding exertion and good performance.

The main components of each of the 14 general effectiveness-enhancing factors are summarized in Table 5.5.

Table 5.5 Components of 14 effectiveness-enhancing factors

Factors	Components
Achievement, orientation, high expectations	• Clear focus on the mastering of basic subjects • High expectations (school level) • High expectations (teacher level) • Records on pupils' achievement
Educational leadership	• General leadership skills • School leader as information provider • Orchestrator of participative decision-making • School leader as coordinator • Meta-controller of classroom processes • Time educational/administrative leadership • Counsellor and quality controller of classroom teachers • Initiator and facilitator of staff professionalization
Consensus and cohesion among staff	• Types and frequency of meetings and consultations • Contents of cooperation • Satisfaction about cooperation • Importance attributed to cooperation • Indicators of successful cooperation
Curriculum quality/opportunity to learn	• The way curricular priorities are set • Choice of methods and text books • Application of methods and textbooks • Opportunity to learn • Satisfaction with the curriculum
School climate	(a) *Orderly atmospheres* • The importance given to an orderly climate • Rules and regulations • Punishment and rewarding • Absenteeism and dropout • Good conduct and behaviour of pupils • Satisfaction with orderly school climate
	(b) *Climate in terms of effectiveness orientation and good internal relationships* • Priorities in an effectiveness-enhancing school climate • Perceptions on effectiveness-enhancing conditions • Relationships between pupils • Relationships between teacher and pupils • Relationships between staff • Relationships: the role of the head teacher • Engagement of pupils • Appraisal of roles and tasks • Job appraisal in terms of facilities, conditions of labour, task load and general satisfaction • Facilities and building
Evaluative potential	• Evaluation emphasis • Monitoring pupils' progress • Use of pupil monitoring systems • School process evaluation • Use of evaluation results • Keeping records on pupils' performance • Satisfaction with evaluation activities

(continued)

Table 5.5 (continued)

Factors	Components
Parental involvement	• Emphasis on parental involvement in school policy • Contacts with parents • Satisfaction with parental involvement
Classroom climate	• Relationships within the classroom • Order • Work attitude • Satisfaction
Effective learning time	• Importance of effective learning • Time • Monitoring of absenteeism • Time at school • Time at classroom level • Classroom management • Homework
Structured instruction	• Importance of structured instruction • Structure of lessons • Preparation of lessons • Direct instruction • Monitoring
Independent learning	No sub-components
Differentiation	• General orientation • Special attention for pupils at risk
Reinforcement and feedback	No sub-components

The range of components within factors in several cases shows that effectiveness-enhancing conditions are measured in terms of:

(a) priorities assigned to factors and components; i.e. attitudes, beliefs, goal statements;
(b) factual state of affairs relevant to factors and components;
(c) appraisal and judgement on the degree to which factors and components are realized.

Particularly with respect to the latter category (appraisal) there is the danger of reactivity in the measurement of (hypothetical) effectiveness-enhancing conditions, because the judgement on processes and antecedent conditions may be coloured by knowledge about outcomes and "dependent variables".

The divergence in choice of elements for instruments across sources (i.e. instruments used in school effectiveness studies and school diagnosis instruments) is somewhat inflated, because there are sometimes rather slight differences between elements. It should also be noted that divergence at item level does not preclude that elements will be correlated and be shown to be subsumable under

common headings, as supported by data analytical procedures like factor analyses. On the other hand it is quite clear that there is little agreement, at the operational level, on the substance of the key factors that are supposed to determine school effectiveness.

A further observation is that most of the factors are broad, in the sense that there is a wide range of components and elements. This is particularly the case for educational leadership and school climate. The broadness of the factors makes it hard to decide which of the set of elements is supposed to be crucial in enhancing effectiveness. Both the divergence and the broadness of the factors make summary review and qualitative research synthesis rather hazardous, because operationalizations of the same general factor may be quite different across studies.

A third and final observation is that the factors are not mutually exclusive. Zones of overlap exist between:

- achievement orientation in policy and climate;
- evaluative potential and monitoring as an aspect of structured teaching;
- curriculum aspects and coordination and consensus;
- educational leadership and use of students' records (also an aspect of evaluative potential);
- participatory decision-making and consensus.

The main conclusion from the analysis of instruments used in school effectiveness research is that there is great divergence among studies, that each project leader appears to be reinventing the wheel in the area of instrument development for measuring effectiveness-enhancing school and classroom variables and that there are few commonly used standardized research instruments to measure factors that are supposed to be the core of effectiveness-enhancing conditions.

More Detailed Definitions and Exemplary Items of the Main Factors

A detailed overview of elements and item examples of the 12 factors is given in Scheerens and Bosker (1997, Chap. 4). The summary Table 5.6 gives an update, which is based on some 40 studies that were carried out in the period between 2000 and 2005 cf. Scheerens et al. (2007).

Table 5.6 Component and sub-items of 14 effectiveness-enhancing factors

Factors	Components	Sub-components and exemplary items
1. Achievement, orientation, high expectations	1.1 Clear focus on the mastering of basic subjects	The emphasis the principal places on the learning of basic skills like reading and science as a contrast to social and creative skills
	1.2 High expectations (school level)	(1) What percent of the students in this school do you expect to complete high school? (2) What percent of the students in this school do you expect to complete a 4-year college degree?
	1.3 High expectations (teacher level)	4 items measuring teacher's proachievement beliefs. Sample items: (1) The teachers in this school believe that learning is important. (2) The teachers at this school really believe that all pupils can achieve. (3) The teachers are only interested in the pupils who do well in tests and exams. (4) The teachers in this school seem to like teaching
	1.4 Records on pupils' achievement	• The school keeps achievement records on all pupils • The school uses achievement records to compare itself with other schools and with earlier performance
2. Educational leadership	2.1 General leadership skills	5 items measuring firm and purposeful head teacher: (1) The head teacher takes action if a teacher's performance is inadequate. (2) The head teacher ensures that teachers are given support to improve their teaching, if they need it. (3) The head teacher encourages staff more than he/she criticizes them. (4) The head teacher makes clear that the quality of teaching and learning at this school are his/her foremost priority
	2.2 School leader as information provider	Degree, timeliness and quality of information provision. The head teacher ensures that there is enough information on the work of colleagues in order to reach sufficient coordination of tasks
	2.3 Orchestrator of participative decision-making	Degree to which principals and teachers shared equally in decision-making process. Sample item: how are decisions made at your school from: student retention policies, use of school funds; supplies and computers, selecting methods (only administrator decides … administrators and teachers decide jointly)

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	2.4 School leader as coordinator	The school leader as an initiator of staff meetings
	2.5 Meta-controller of classroom processes	9 items measuring integrated leadership: e.g. (a) there is evidence of significant instructional leadership in the school, (b) significant instructional leadership comes from a principal or other school-based administrator, and (c) significant instructional leadership comes from teacher or group of teachers. (d) The actual influence of teachers over curriculum (e) the actual influence of teachers over instruction (f) the actual influence of teachers over student assessment (g) the actual influence of principals over curriculum (h) the actual influence of principals over instruction, and (i) the actual influence of principals over student assessment
	2.6 Time educational/administrative leadership	• The number of hours a head teacher teaches • Total number of hours for managerial, non-teaching activities • Division of school leader activities over administrative/organizational, instructional leadership, contacts with parents, own professional development • The number of times per year/month a head teacher attends lessons, discusses pupils" functioning with teachers • Teachers are content with the relative emphasis the head teacher spends on instructional versus other leadership tasks • The degree to which teachers are satisfied with stimulating effectiveness-enhancing leadership
	2.7 Counsellor and quality controller of classroom teachers	The concept is measured by three components: principal observation of classes (frequency p/y); evaluation of teachers (scale 1–3); evaluation of school quality by principal (scale 1–3)
	2.8 Initiator and facilitator of staff professionalization	• The head teacher encourages further education of teachers in a selective, targeted way

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
3. Consensus and cohesion among staff	3.1 Types and frequency of meetings and consultations	• Number of formal staff meetings with the head teacher • Frequency of informal meetings among groups of teachers • Informal contacts between staff
	3.2 Contents of cooperation	The degree of consistent practice among teaching staff indicated by within school standard deviation; one scale consisting of 8 items measuring the extent to which teachers' testing and grading practices are regulated (by their fellow teachers). Items concern assignment of report grades, time of announcing tests, content of tests, grading of tests, frequency of tests, discussing test results with class, frequency of quizzes. Five-point scale (1 = no rules—5 formal rules exist and are influential: items concern textbook, content to be covered, learning goals to be achieved, sequence of topics, amount of time spent on topics, homework, assignments to be made, teaching methods to be used. Five-point Likert scale (1 = no rules—5 formal rules exist and are influential)
	3.3 Satisfaction about cooperation	• Satisfaction in relation to colleagues with respect to allocation of duties and coordination concerning: – Variety in interests – Professional competence – Supporting school improvement – Involvement in pupils' learning and satisfaction – The amount of curriculum/ 'techniques' – Discussion in team meetings – Acceptance, support and opportunity to cooperate – Cooperation at school and within the team
	3.4 Importance attributed to cooperation	Measured by three scales: time for teacher collaboration (time), improving instruction through discussion (innovation), encouragement of encouragement in teacher participation (participate)

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	3.5 Indicators of successful cooperation	4 items asking the teachers and head teacher on the school's aims and values and how these are implemented through teaching and learning. Sample items: (1) Teachers at this school are all committed to the school's aims and values. (2) Teachers follow the same set of rules about pupil behaviour. (3) There is general agreement amongst the teachers about what are effective teaching. (4) Teachers and the head teacher agree on how teachers and pupils should behave towards each other
		The degree to which teachers share similar goals and beliefs. Sample items: Most of colleagues share my beliefs and values about what the central mission of the school should be. Goals and priorities for the school are clear (strongly disagree…strongly agree)
4. Curriculum quality/ opportunity to learn	4.1 The way curricular priorities are set	Scale of 3 items measuring: the principal watches over the implementation of subject curricula (range 1–9) • The extent to which subject matter provision is determined (i.e. guidelines are developed) by the ministry, the school board, the school team • Knowledge about core objectives arithmetic/math and science, the school work plan • The importance of a good range of extra-curricular activities for the school's effectiveness
	4.2 Choice of methods and text books	• Availability of books for language and math • Well-functioning methods for spelling, decoding, reading comprehension, composition writing and math, meaning: – A clear line with regard to subject matter content – Clear directives for instruction and testing – A step-by-step approach for the low achievers – A clear distribution of minimum competency goals over school years • Which language methods (in which group) • Which arithmetic-math methods (in which group) • Method for science

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	4.3 Application of methods and text books	• Knowledge of the manual for arithmetic/math/science methods • The time the method is being used • Considering transfer to other methods • Which part and which chapter in the beginning of the school year • Which part and which chapter now • Keeping sequence in the method • % of subject matter dealt with at the end of the school year • Progress in method at the end of the school year • Other material for arithmetic/math/language/science than prescribed in method • Use of a calculator • % of pupils being in a position to use a calculator
	4.4 Opportunity to learn	• % of time for arithmetic/math/science spent on method • Division of lessons to subject matter components • Other subject matter areas (within the subject) • Number of lessons per subject matter area • Which test items link up with education taught so far (for arithmetic/math and science)
	4.5 Satisfaction with the curriculum	• The extent of satisfaction with the curriculum now and 5 years ago • Satisfaction with the curriculum and the teaching materials • Satisfaction with the choice of subjects offered • Effectiveness of the curriculum's coordination within in the school • Successes with respect to extra-curricular activities and curriculum development over the past 5 years • The degree to which the work at school is considering interesting • The extent to which a curriculum is modern • Lessons: – Number of lessons that stir the imagination – Diversity of subjects

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
5. School climate	(a) *Orderly atmospheres*	
	5.1 The importance given to an orderly climate	Scale of 7 items measuring: teachers are keen on a pretty and tidy classroom, students are encouraged to act orderly (range 1–21)
	5.2 Rules and regulations	One scale consisting of 4 items measuring whether teacher made rules exist to control student behaviour. Items concern student being late, truancy, classroom disruption, cheating at tests. Five-point scale (1 = no rules— 5 = formal rules exist and influence my work)
	5.3 Punishment and rewarding	• % of pupils being disciplinary punished last year • Number of rewards mentioned by the school head • Number of punishments mentioned by the school head • Rewards/punishments ratio • Teacher rewards work more than punishment • Teacher rewards behaviour more than punishment • Forms of rewards by school head (a.o. praise) • Forms of punishments by school head (a.o. verbal warnings, confinement) • A clearly applied system of punishment and rewarding at the school
	5.4 Absenteeism and drop out	School social problems scale includes 22 items that tap principal ratings of criminal activity, attendance problems, high-risk activities, and school-level parental investment in student health and well-being (high scores = more problems)
	5.5 Good conduct and behaviour of pupils	• Other pupils do not encourage a child teasing another child • Teachers and pupils see to it that teaching–learning processes are undisturbed • Teachers create a learning environment in which pupils can work in a task-oriented way • See to it that nobody disturbs a teacher during the lesson • The pupils behave well when the teacher leaves the classroom

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• The lessons are not often disturbed by noise down the hall • Level of pupil sound in the classroom • Level of pupil movement in the classroom • Teachers' audibility in the classroom • Pupils' behaviour around the school • Strengthening pupils' behaviour • The level of unaccepted pupils' behaviour now and 5 years ago • Important successes and problems with respect to pupils' behaviour and discipline now and 5 years ago • The school's high standards of pupil behaviour • The frequency school heads or team are being confronted with the following behaviour (of grade 6) – Vandalism – Theft
	5.6 Satisfaction with orderly school climate	• Satisfaction with respect to safety at school, behaviour in the classroom, the school and teachers being attentive • Satisfaction with respect to pupils' behaviour • Degree of satisfaction with pupils' behaviour now and 5 years ago • The extent to which teachers set an example in their behaviour to pupils • Satisfaction with respect to precautions/the way the school handles vandalism, drugs, alcohol and tobacco
	(b) *Climate in terms of effectiveness orientation and good internal relationships*	
	5.7 Priorities in an effectiveness-enhancing school climate	• Effectiveness-enhancing conditions for a school – A caring pastoral environment – Positive interpersonal relationships for staff and – Students – The encouragement of a positive attitude to school (pride in school) – Shared goals and values by staff and students – High level of pupil motivation – Students satisfaction • Effectiveness-enhancing conditions for your school

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		– Students feel valued as people – Encouragement of student responsibility
	5.8 Perceptions on effectiveness-enhancing conditions	• Effectiveness-enhancing conditions of a school: – Teacher motivation – Teacher commitment/effort – Personal effectiveness of teaching staff – Commitment/enthusiasm of teaching staff • Effectiveness restricting conditions of a school: – Heavy workload – Low staff morale – Lack of commitment and enthusiasm by some staff – High teaching staff absence rates
	5.9 Relationships between pupils	• How do you feel about relationships between pupils • Communication between pupils • Pupils want to belong to the school and to each other
	5.10 Relationships between teacher and pupils	• How do you feel about relationships between pupils and teachers • Contacts with pupils are open and pleasant • The teacher/pupil social relations are good • The team tries to understand pupils' needs • Communication with teachers • Teachers like pupils, support them, want them to associate nicely, know what every pupil wants, treat them fair, etc. • Did the school have success with respect to better relationships between teachers and pupils the past 5 years • Team functioning with respect to controlling pupils (firm but friendly relations)
	5.11 Relationships between staff	Four items measuring warm staff atmosphere. Sample items: (1) The teachers at this school are friendly towards each other. (2) The teachers work well together. (3) If I have a problem, I will get support from other

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		teachers. (4) Newly qualified teachers at this school are supported by experienced teachers
	5.12 Relationships: the role of the head teacher	• Communication between head teacher and pupils • Head teacher listens to ideas/opinions/complaints from pupils about the climate and atmosphere) • Relationships between school head and teachers • The school head: – Trusts his team members – Can easily be approached – Progresses job satisfaction – Takes suggestions and ideas of teachers with respect to work climate and – Sphere serious – Pays attention to solving/improving mutual relations in case of conflicts • The behaviour of school head evokes conflict
	5.13 Engagement of pupils	• Pupils have a say in what happens at school • Pupils co-decide about what happens at school • Pupils are proud of the school and show responsibility • Did the school have success with respect to pupils' responsibility the past 5 years
	5.14 Appraisal of roles and tasks	• Teaching/other tasks • Role clarity (clearly described tasks) • Job variety • Degree of job satisfaction
	5.15 Job appraisal in terms of facilities, conditions of labour, task load and general satisfaction	*Job appraisal in terms of facilities, conditions of labour, task load and general satisfaction* • Sufficient facilities (methods/materials) to efficiently carry out work • Salary and (secondary) conditions of labour • Competent authority passing onto a rewarding system based at personal commitment and motivation of teachers • Importance of part-time appointments • Opportunities for career enhancement

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• Task load (general anticipatory and perceived psychosocial mental strain): – In general – Own task load • Satisfaction with respect to working hours • Teachers believe they are overworked and under pressure • Average absenteeism of team members now and 5 years ago • Quality of working life • Satisfaction with respect to working with pupils • Enthusiasm for the work/the school (now and 5 years ago) • Attention for extra-curricular activities • Feeling valued in functioning as a teacher • Opinion with respect to teachers' motivation • Successes/problems with respect to teachers' motivation during the past 5 years
	5.16 Facilities and building	• Classrooms/school/school building/playground clean, neat and well equipped • Sufficient space in/around the school • Sufficiently good facilities in and around the school • No problems with respect to the school's entrance and with respect to stairs and halls in the school • Service quality in the area of safety, advice, care, health and canteen/stay-over facilities
6. Evaluative potential	6.1 Evaluation emphasis	Measured whether or not there is a monitoring system, whether summative assessment is employed and whether there is a central registration of pupils' achievement
		Scale of 9 items measuring "Evaluation policy" variable on school level: school wide use of tests for basic subjects, team evaluation of student progress, standardization of achievement test procedures for basic subjects

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	6.2 Monitoring pupils' progress	• A strong emphasis on the evaluation of test results • Agreements and/or rules at school level with respect to testing/registration • At our school pupils' progress is regularly tested/we handle a good testing system for progress registration to register problems with pupils in time and to take appropriate measures • The extent to which a department head evaluates the learning progress in the department • In groups 1 and 2 attention is paid to early signalizing so-called "pupils at risk" with regard to speech-language, social–emotional, auditive, visual-spatial and motor development, concern for more cognitive activities and the task and work attitude • The extent to which reading and arithmetic are tested • Evaluation of pupils' progress takes place by means of standardized progress tests • What is pupils' assessment based on (national standards, comparison with other schools, progress of the child itself) • Does the school handle achievement standards for individual pupils/standards at school level • (Written) rules for promotion to the next year/retention yes/no • Decision on promotion/retention based on opinion teacher • Is the school posted on pupils' functioning in further education
	6.3 Use of pupil monitoring systems	• Pupils' progress being administered in a pupil monitoring system at school level • Evaluating pupils' progress in basic skills at least twice a year by means of a pupil monitoring system • Registration of pupils' progress in individual pupil files, in group surveys, in central pupil monitoring system • Which pupil monitoring system is being used and do all teachers use the same pupil monitoring system

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	6.4 School process evaluation	• Has the school been assessed during the past 5 years by means of an instrument for school self-evaluation • Which aspects are structural tested/evaluated, analysed and, if necessary, improved: – Pupil satisfaction – Teacher achievement on the basis of pupil data – Teacher satisfaction on the basis of … – Functioning of the school management – Resource expenditure – Courses and teaching – Provision of education – New teaching methods – Dissemination of innovations – The process of educational improvement – Implemented changes – Policy formation
	6.5 Use of evaluation results	• The school being aware of possible level of changes in pupil performance during the past 5 years • The school being aware of its position with respect to pupil performance with regard to other schools having a comparable pupil population • For how many subjects is it possible to compare the present average achievement level to 5 years ago • For how many subjects does the school compare pupil progress with other schools • Discussing pupils' progress and development regularly and systematically • Evaluation of pupil performance: – Leads to adjustment of instruction and learning strategies – Supports assignment to ability groups – Changes in teaching strategies • Comparisons in achievement are being used for educational improvement • Using former pupil data for educational improvement

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	6.6 Keeping records on pupils' performance	• Is keeping records on pupils' performance dealt with in the school work plan • If yes, indications for keeping records on pupils' performance concern the recording of it • Teachers keep records on pupils' development and progress • Does the teacher keep records on language progress • Total number of registrations by teacher • How often does keeping records on individual pupil's progress in documents open to the school head occur • Method of registration of learning progress: (a) Standardized data (b) Judgement by individual teacher (c) Both a and b (d) There is no registration • Registration school progress: – Not – In individual pupil file – In group summary – In central pupil monitoring system • Are pupils' data kept up with through the entire school career • If yes, by means of automatized computer system • Frequency in which summaries of registration data are presented: – Per pupil – Per teacher • Group summaries of pupils' achievement are made • Use summaries per pupil/teacher for … • Record results written assignment • Record test results • Execute an error analysis • Process pupils' achievement in pupil monitoring system at school level • Frequency of written reports to parents (per school year/group)

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• Quality of reporting of pupils' progress (all-embracing, exploratory and valuable information on pupil's progress) • The school pays a lot of attention to reporting towards pupils and parents • Written pupils' report when pupils pass to next school year
	6.7 Satisfaction with evaluation activities	• The degree of satisfaction with the student assessment/monitoring system now and 5 years ago • During the past 5 years, did the school succeed in establishing: – Improved record-keeping/student profiles – Improved monitoring of pupils' progress • The team's satisfaction with respect to the amount of attention paid to improving education
7. Parental involvement	7.1 Emphasis on parental involvement in school policy	• Strong parental support as an important condition for school effectiveness • Little parental support impedes effectiveness • School heads and teachers are open for suggestions from parents • The school emphasizing the importance of parental involvement with respect to education and pedagogical affairs • The school being open for parents attending lessons • The school has a parents' association of which parents can become a member on a voluntary base • Are parents in parents' committees, parents' councils or participation councils reflecting the pupils' population and is this aimed for • Agreements with respect to home visits • Facilities for parents to be present in the school • Parents' complaints are taken seriously • Agreement with the following pronouncements: – Parental involvement is considered positive – Parents are allowed to influence education's organizational structure

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		– Parents are allowed to influence educational contents – The school's and parents' responsibilities should be clearly defined – Disappointing achievement is often due to parents not supporting the school • A parent activity programme is drafted yearly • The school stimulates that as many parents as possible attend the individual talks about their child's progress • The school pays specific attention to parents who are hard to reach
	7.2 Contacts with parents	"Parental Involvement" was created by averaging the responses to three items: whether the parent participates in school-related activities (1 = seldom, 2 = sometimes; 3 = always); whether the parent knew his/her child's teacher meetings (1 = never, 2 = a little, 3 = a lot), and whether the parent attended parent–teacher meetings (1 = never or seldom, 2 = almost always, 3 = always)
	7.3 Satisfaction with parental involvement	Parents' self-efficacy beliefs for their children's reading achievement were measured using an 18-item questionnaire by asking both parents perspectives, 10 items measuring parents' belief that they have the competence to successfully teach their children. Sample item: (1) by reading to my child, I can help my child become a better reader. Eight items measuring parents' attributions for their children's success or failure. For example, children are good readers because they have a natural ability
8. Classroom climate	8.1 Relationships within the classroom	Positive interaction with teacher: Likert scale based on frequency of 4 items; ranges from 0 (low) to 3 (high). Items: (1) Have you been told that your work is good? (2) Have you been asked questions in class? (3) Have you been praised for answering difficult question correctly? (4) Have you been praised because your written work is well done?

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		Negative interaction with teacher: Likert scale based on frequency of 5 items; values range from 0 (low) to 3 (high) sample items: (1) have you been given out to because your work is untidy or not done on time? (2) Have you wanted to ask or answer questions in class but were ignored? (3) Have you been given out to for misbehaving in class? (4) Teachers pay more attention in class to what some pupils say than to others. (5) I find most teachers hard to talk to
	8.2 Order	• Fairness/firmness (control in the classroom) • Classroom scores on: – Order in the classroom • Rules in the group are clear for each pupil • Creation of an orderly, quiet work environment • Situation with respect to control (firm but friendly relations) on pupils now and 5 years ago
	8.3 Work attitude	Inattentive behaviours in the classroom were measured as dimensional, weighted composites of 4 items: (1) Cannot concentrate on any task; easily distracted (2) Lacks perseverance; is impatient with difficult or challenging tasks. (3) Easily frustrated; short attention span (4) Aimless; impulsive activity
	8.4 Satisfaction	• Classroom fun factor The fun factor is to give an indication of whether or not it was an enjoyable experience to be a pupil in a particular teacher's class. the 'fun factor' is the sum of all 'yes' responses to the 8 items that follow: – Did the teacher smile often – Was there positive physical contact with pupils – Did the teacher show a sympathetic interest in the children other than as learners – Did the teacher chat to the pupils about non-work matters on any occasion during the day (whether pupil or teacher initiated)

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		– Was communication between children generally cheerful – Was the children's behaviour generally relaxed – Were there any jokes and/or was there any laughter in which the teacher was involved? (this does not include jokes at the expense of other pupils) – Was there any sign that pupils wanted to be in the classroom outside of class teaching time, either before or after sessions
9. Effective learning time	9.1 Importance of effective learning	• Emphasis on – Developing better policy and better procedures to enlarge instruction time • Impeding/progressing school effectiveness: – Good registration of presence and absenteeism – Good class management – Give high priority to homework
	9.2 Time	Scale of 6 items measuring: starting lessons on time, prevention of disturbances, rules on student truancy (range 1–18)
	9.3 Monitoring of absenteeism	• % of pupils truanting • The way the school handles absenteeism and lateness • Satisfaction with respect to pupils' presence now and 5 years ago
	9.4 Time at school	• Number of school days • Number of teaching days/hours – Number of teaching days per school year – Number of full teaching days per school week – Number of semi-teaching days per school week – Total number of hours per school week – Length of a school day • % of cancelling of lessons • Number of days with no lessons due to structural causes • % of total number of hours indicated on the table

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• Measures to restrict cancelling of lessons as much as possible • Policy with respect to unexpected absenteeism of a teacher • (In school work plan) agreements on substituting teachers
	9.5 Time at classroom level	• Number of lessons on timetable per school year • A lesson consists of how many minutes • Amount of teaching hours for language/arithmetic • Amount of minutes for arithmetic/physics per week • Duration last arithmetic lesson in minutes • Accuracy with respect to starting and finishing lessons in time now and 5 years ago • Number of lessons that are cancelled • Satisfaction with respect to available amount of time for working in the classroom
	9.6 Classroom management	• Attention for classroom management in the school work plan – With respect to lesson preparation – Rules and procedures for the lesson's course • Situation with respect to task orientation during lessons (now and 5 years ago) • Average % of teachers spending time on: – Organization of the lesson – Conversation (small talk) – Interaction with respect to the work – Supervision (pupil activities/behaviour) – Feedback/acknowledgement • Average time during lesson spent on discussing homework, explaining new subject matter, maintaining order • Sources of loss of time during lessons: – Pupils do not know where to find equipment – Disturbances due to bad behaviour of pupils – Frequent interruptions – Loss of time due to lengthy transitions from one activity to the next

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		– Unnecessary alterations in seating arrangements – Frequent temporarily absence of pupils during lessons – Waiting time for individual guidance – Many (more than 3) teacher interventions to keep order – Lack of control on pupils' task-related work
	9.7 Homework	• Attention for assigning homework at school/agreements in school work plan • Homework after last (arithmetic) lesson: yes/no • Number of homework assignments per week • Type of homework (arithmetic/language) (reading/composition writing) • Amount of homework • Amount of time needed for homework (per day) • Extra homework for low-achieving pupils • Successes and problems now and 5 years ago with respect to: – Prioritizing homework – A consistent homework policy • Whether homework assignments are graded or not
10. Structured instruction	10.1 Importance of structured instruction	• Emphasis in school's policy on – The quality of teaching – Encouraging pupils to take responsibility for their own learning process – Teacher-independent learning – Emphasizing exam preparation – Sufficient "challenge" for both high and low-achieving pupils • To what extent agreed upon: – Whole class instruction gives the best results – Discovery learning mainly needs to happen outside the school – Pupils acquire less knowledge when different pupils do different tasks – Repeating a year often benefits pupils' development

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		– The high-achieving pupil is especially the victim of individualized education – Individualized education benefits all pupils – When dividing pupils into groups achievement will do as criterion
	10.2 Structure of lessons	• Direct instruction divided in: – Looking back daily – Presenting subject matter – Guided practice – Giving feedback and correction – Independent practice – Looking back weekly/monthly • Teacher uses a lesson plan
	10.3 Preparation of lessons	• Lesson preparation building upon: – Lessons formerly taught – Written plan – Other teachers/math specialists – Text books – Standardized tests • Most important information source for planning arithmetic/math lessons (lesson content, way of presentation, homework, tests) – Core objectives – School work plan – Manual – Text book – Other source books • The subject matter is the central factor when teaching
	10.4 Direct instruction	• Attention for instruction in the school work plan • Indications in school work plan with respect to: – Clear objectives of instruction – Construction of the instruction – Way of presenting subject matter – The use of instructional materials • Explanation or help to individual/groups of pupils in or outside the lesson • Teachers deal with subject matter that corresponds to the lesson's aim • Teacher explains at the beginning of the lesson to what prior knowledge the subject matter corresponds • Teacher gives pupils the chance to raise questions about the last lesson

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• Teacher explains beforehand what pupils have to know at the end of the lesson • Teacher knows what to achieve with the lesson • Lesson objectives are clear to pupils • Teacher applies instructional methods to increase pupil's achievement • Teacher deals with only one subject matter component at the time • Explanation in small successive steps • Teacher takes next step when preceding step is understood • Teacher gives concrete examples • It appears from pupils' reactions that the teacher explains the subject matter clearly • Teacher poses intellectual questions that invite pupils to participate actively • After posing a question the teacher waits to let the pupils think • Teacher gives many pupils a turn • A lot of interaction between teacher and pupils • Pupils respond well to questions posed by the teacher • Teacher have pupils practised under guidance • Teacher continues until all pupils have mastered the subject matter • Explanation is clear • Teacher involves pupils in instruction • Teacher takes care that pupils are concentrated during instruction • During instruction immediate feedback to answers of pupils • The lesson displays a clear structure • At the end of instruction summary of subject matter (by teacher/pupils) • Pupils get tasks they can handle • Group work, if appropriate • Teacher's activities (controlling) when pupils work on assignments • Teachers take time to help pupils with tasks • Pupils know which tasks are to be carried out • Teacher sees to it that pupils work in a concentrated way during assignments

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• Teacher sees to it that pupils work task-oriented during assignments • From pupils' reactions it appears that everyone knows what he or she has to do • There is sufficient control on pupils doing the assignments they are supposed to do • Pupils work at a good pace • % of time during lessons in which assignments are discussed • Analysis of mistakes • Checks on homework
	10.5 Monitoring	• Is monitoring of pupils' achievement mentioned in the school work plan • Indications concerning: – Pupils' written assignment – The use of tests • % of lessons containing tests • The number of tests, hearings • Types of tests per school year (a.o. posing questions in class, own tests, curriculum-embedded tests) • Which procedures are used to assess pupils' achievement with respect to arithmetic • Progress in pupil learning outcomes is measured by means of (curriculum-embedded) tests • Teacher uses checklist for oral hearing of pupils • The way the teacher prepares pupils for tests • Teacher checks whether all pupils have reached the minimum goals • Teacher checks up on difference between expected and actual pupil achievement • Compare pupil achievement to: – Former pupil achievement – Fellow-pupil achievement – Norms and standards • In what way is arithmetic/math work of a pupil judged (absolute criterion, class average, etc.) • Are test results used for individual help, extra explanation

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		• Taking action in connection with test results • Use learning progress for: – Preparing a program for individual pupil – Reporting to parents – Informing teacher about next group – Evaluating the school's functioning – Putting pupils into (parallel) classes – Selecting pupils for teaching programmes (enrichment/remediation) – Grouping pupils within classes – Other • The degree of pupils' progress has an effect on class level (e.g. other grouping patterns, more or less instruction, etc.) • Successes/problems with respect to preparation for tests over the past 5 years • Review and correct written assignment of pupils • Use of curriculum-embedded tests • Use of curriculum-independent tests • Use of self-made tests
11. Independent learning	No sub-components	• Attention for independent learning in school work plan • Teacher-independent learning is being encouraged yes/no • If yes, indications concerning: – Relation instruction/processing time – Organization of independent learning – Other types of differentiation • State of affairs with respect to teacher-independent learning/independent learning • The extent to which pupils are responsible for their own work • The extent to which pupils are responsible for their own work during a longer period • The extent to which pupils are able to choose their own assignments • The extent to which pupils' cooperation is encouraged by teachers • In case of independent learning, do pupils work: – On the same subject

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		– On various subjects per group of same level – On the same subject at own level – On various subjects at various levels • Opportunity for pupils to plan the school day themselves • Successes and problems with respect to teacher-independent learning/independent learning
12. Differentiation	12.1 General orientation	The extent of streaming and associated curricular differentiation in the school; scale ranging from 0 (mixed ability base classes) to 4 (highly streamed) • How to deal with differences between pupils in arithmetic/math attainment levels during lessons (all pupils the same subject matter, …) • % of lessons in which pupils: – Work on the same subject – Work on two subjects – Work on three or more subjects • How often do pupils work individually or in pairs • % of teacher time spent on communication with the class, groups and individuals • Criteria with respect to subject matter provision/grouping: – Achievement – Results standardized test – Results diagnostic test – Results oral test – Teachers' recommendations – Parents' wishes – Pupils' wishes – Method's demands • Pupil grouping within the class: – No grouping – Age groups – Level groups – Interest groups – Other • Frequency of regrouping pupils (evt. of more classes) on behalf of level groups • Problems and successes with respect to differentiation in the past 5 years • Subject matter mastery adapted to slow and fast learners

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
	12.2 Special attention for pupils at risk	• Policy with regard to low-achieving pupils • School policy is explicitly aimed at catering for a wide range of educational needs: in other words, clear directives and structural attention for pupils with problems • Catering for special individual educational needs concerning: – Diagnosing pupils "at risk" – Remedial teaching – Cooperation with special education – Drafting intervention plans – Drafting group plans • Amount of extra time teachers are prepared to spend on problem pupils • Extra provisions for problem pupils • Low-achieving pupils get more time for reflection, extra attention, instruction, help, material and exercise material • Provisions/approved methods for preventing (teaching) problems • Check systematically which subject matter is not being mastered • Group teachers having expertise with regard to diagnostic test administration • Group teachers are able to translate test data into intervention plans
13. Reinforcement and feedback	No sub-components	*Reinforcement* • Is feedback in connection with pupils' achievement discussed in the school work plan • Indications for feedback in connection with pupils' achievement are related to discussion by the teacher • How often, in arithmetic/math lessons, do you take the following action when pupils answer wrongly (a.o. correct wrong answer, pose different question) • During the lesson feedback is given and pupils' mistakes are corrected • When pupils carried out an assignment it is discussed immediately • The teacher explains what was wrong when he returns the tests • Teacher gives pupil as much as possible real and positive feedback to achieved results • Frequency of discussing learning progress with pupils • Low-achieving pupils get extra feedback

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		Feedback • Results written assignment is discussed with pupil if necessary • Results curriculum-embedded test are discussed with pupil if necessary • Results method-independent tests are discussed with pupil if necessary • Results of self-made tests are discussed with pupil if necessary • A differentiated supply based on tests is offered • Quality/suitability of feedback • State of affairs with respect to giving constructive feedback now and 5 years ago • Problems with respect to inadequate feedback
		Two items measuring positive reinforcement (teachers): (1) The school has a system for rewarding pupils who work hard and/or make good progress even if they do not get high grades. (2) A pupil who works hard or makes good progress is noticed and praised
		Four items measuring monitoring and rewards: (1) I am set targets for my learning by my teachers which are individual to me and not for the whole class. (2) The school has rewards for pupils who work hard or make good progress even if they do not get high grades. (3) A pupil who works hard or makes good progress is noticed and praised. (4) Teachers notice those pupils who are not working as well as they could and try to make them work harder
		Measures the quality of teaching. Example: the frequency of monitoring whether pupils have mastered the learning content, the frequency of repeating learning content where necessary, the frequency of evaluating pupils' progress and giving help to pupils as needed
		Measuring the degree of regular feedback given to pupils on achievement tests

(continued)

Table 5.6 (continued)

Factors	Components	Sub-components and exemplary items
		Scale measuring teacher feedback. Items use from the School Organisational Health Questionnaire (SOHQ) (teachers perceptions)

References

Cotton, K. (1995). *Effective schooling practices: A research synthesis*. 1995 Update. *School Improvement Research Series*. Northwest Regional Educational Laboratory.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Hendriks, M. A., & Scheerens, J. (2013). School leadership effects revisited: a review of empirical studies guided by indirect-effect models. *School Leadership and Management, 33*(4), 373–394.

Levine, D. K., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, Wise: National Centre for Effective Schools Research and Development.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J. (1987). *Het evaluerend vermogen van onderwijsorganisaties*. (The evaluative potential of educational organizations) Universiteit Twente, Toegepaste Onderwijskunde, oratie.

Scheerens, J. (1989). Process indicators of school functioning. *School Effectiveness and School Improvement, 1*(1), 61–80.

Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. *Interantional Journal of Educational Research, 13*(7), 691–706.

Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.

Scheerens, J. (Ed.). (2012). *School leadership effects revisited. Review and meta-analysis of empirical studies.* Dordrecht, Heidelberg, London, New York: Springer.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). Review and meta-analyses of school and teaching effectiveness. *Enschede: Department of Educational Organisation and Management, University of Twente*.

Stringfield, S. C., & Slavin, R. (1992). *Raising societal demands, high reliabilty organizations, school effectiveness, "Success for All", and a set of modest proposals*. Address to the ICO-symposium, Enschede.

Chapter 6
System-Level Context and Effectiveness Enhancing Policies

Keywords System effectiveness · Functional decentralization · School autonomy · The societal context of education systems · Accountability (administrative · Market-oriented and professional) · The structure of secondary school systems · Comprehensive school systems · Categorical school systems · Equity oriented policies · Initial teacher training · Quality oriented educational programs

Introduction

This chapter completes the modelling of educational effectiveness by addressing system-level structures and policies. The distinction between context and ecology and policy amenable factors associated with educational attainment is used, once again, as the ordering framework for describing the most relevant conditions. First of all, the larger societal context is mentioned as it interacts with the educational (sub-) system. On the one hand, high expectations exist about the societal benefits of high-quality education; on the other hand cultural, social and economic conditions are likely to affect the performance of the educational system. At the national system level, the distinction between ecology on the one hand and deliberate policies and national programs on the other is not very sharp. The distinction is particularly difficult when reforms have a strong structural component, as in the case of decentralization policies and changing the secondary school system from high stratification to more integrated and comprehensive. Reforms in the domain of accountability, evaluation and assessment are also likely to involve structural adaptations next to procedural facets. Managing the education budget, quality and equity-oriented policies and teacher training are described as examples of malleable policies and inputs at system level.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_6

The Larger Societal Context

The education sector in a society could be seen as embedded in the larger context of economic, social, political and cultural institutions. Salient conditions of this larger context are the general affluence of the country or region, the level of disparity in income of the population, the heterogeneity of the population and cultural norms that could be seen as relevant for the appreciation of education and the motivation of the population to make efforts to do well in education. The linking mechanisms between these aspects of the larger societal context and the functioning of national/regional educational systems are generally not very well specified. Correlations between the economic affluence of a region, for example, and educational performance are sometimes interpreted as effects of education on society (OECD 2010) but might as well be seen as environmental conditions that facilitate or constrain educational performance (Baumert et al. 2005). Here, we will depart from this latter interpretation. Baumert et al. suggest that lack of economic affluence, characterized by relatively large unemployment quota and a relatively large percentage of the working population that depend on social support, could lead to "shared helplessness" in a community and to lack of stimulating examples and role models for young people. The degree to which societies live with inequalities and may have institutionalized them in casts and subcultures could be seen as background to acceptance and institutionalization of inequalities in the education system, while, alternately, very equalitarian societies would be expected to have more equalitarian school systems as well. The Gini index[1] can be used as an overall indicator of the inequality of a nation or region. Obviously, demographic conditions make it easier for some countries than for others to restrict disparities in educational performance within acceptable bands than for other countries. The heterogeneity of the population, particularly in the case of large proportions of first generation immigrants from other cultures, might be expected to influence both performance levels and disparities in outcomes. Appreciation and valuing of education in a country might also be traced back to cultural and religious traditions; this issue, however, is not taken up in this chapter apart from mentioning the operational condition of appreciation and valuing of education as a relevant contextual condition at system level.

Of course policy will also be shaped by party political ideologies and liberal, socialist or conservative domination in governing coalitions; this, for example, is likely to affect the emphasis that is given to equity versus excellence in educational policy.

[1]It measures the extent to which the distribution of income (or consumption) among individuals or households within a country deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. A value of 0 represents perfect equality, a value of 100 perfect inequality" (Source World Bank 2003).

The Education System's Ecology, Areas of Structural Reform

At the level of national and regional educational systems, structural arrangements that are partly given and historically determined, but for another part are the results of long term, processes of educational reform, would fit a category of conditions that at the level of schools and classrooms were described as ecological conditions. In analogy one could speak of the educational system's ecology. Three areas of structural reform have stood out during the past two decades in industrialized, as well as developing countries, these are: *functional decentralization*, the *institutionalization of evaluation and accountability* arrangements and the degree of *structural integration versus segregation* of secondary education.

Functional Decentralization

Expectations about the effectiveness of decentralization policies, to some extent, equal the argument for private schools, namely that autonomous schools are more sensitive to the demands of the consumers of education, and would therefore become more result-oriented and less bureaucratic. Research results generally do not bear out these expectations however (Maslowski et al. 2007). Before reviewing these research results, a more nuanced conceptualization of centralization and decentralization policies is provided by introducing the concept of functional decentralization.

Functional decentralization takes into consideration that systems can decentralize in certain educational functional domains, while being centralized in others. In this way, the issue of centralization and decentralization is dealt with in a more nuanced way, in which "patterns" rather than one dimensional classifications result. The idea of functional decentralization is best illustrated by describing the OECD/INES locus of decision-making instrument (van Amelsvoort and Scheerens 1997).

The procedure used in the OECD education indicator project (INES) to measure "locus of decision making" distinguishes *three* facets of the rather crude distinction between centralisation and decentralisation.

- The tier or administrative level where a decision is taken; this dimension was referred to as the *locus of decision-making*;
- The amount of discretion, or the degree of autonomy of decision-making at a particular administrative level; this facet was called the *mode of decision-making*;
- The particular element of educational administration a decision belonged to; this facet was referred to as the *domain of decision-making*.

These three facets can be related to existing categorizations in the relevant literature, although the use of central concepts is by no means consistent among

authors and publications. The three-dimensional conceptualisation is compared to the terminology as clarified by Bray (1994, p. 819) in an analysis of alternative meanings of centralisation and decentralisation.

The distinction between levels confirms to the concept of *territorial decentralisation*, defined as "the distribution of powers between different tiers of government".

Degrees of autonomy in decision-making at a particular level are reflected in terms that refer to an increase in discretion. Again following Bray, *deconcentration*, *delegation* and *devolution* are modes of decision-making in which an increased amount of decision-making authority resides at a lower level.

"*Deconcentration* is the process through which a central authority establishes field units, staffing them with its own officers".

"*Delegation* implies a stronger degree of decision-making at the lower level. However, powers in a delegated system still basically rest with the central authority, which has chosen to 'lend' them to a local one".

"*Devolution* is the most extreme form of decentralization. Powers are formally held by local bodies, which do not need to seek approval for their actions" (ibid., p. 819).

In the operationalization of this continuum of increasing autonomy, these abstract definitions were avoided and respondents were asked to indicate whether decisions could be taken within the framework determined by a higher level, in consultation with a higher level or in full autonomy.

In order to determine elements or *domains* of educational administration, many categorization schemes are available in the literature (e.g. James 1994; Winkler 1989; Bacharach et al. 1990; Rideout and Ural 1993). The common cores of these categorizations are three main areas

(a) an educational domain (goals, methods, curricula and evaluation procedures);
(b) an organizational, managerial and administrative domain (including human resource management, groupings and assignment and foundational regulations);
(c) a dimension concerning finance and the way financial resources are applied.

A broad range of studies has been conducted to investigate the impact of decentralization and school autonomy on student performance: Caldwell and Spinks (1988), Winkler and Gehrsberg (2000), Whitty et al. (1998), Malen et al. (1990), Summers and Johnson (1996), Wang and Walberg (2001), Woessmann (2001), Fuchs and Woessmann (2004), Mons (2004), OECD (2005) and Maslowski et al. (2007). Some studies have explicitly addressed the effects of school-based management: Leithwood and Menzies (1998a, b), Murphy and Beck (1995), Fullan and Watson (2000) and Bryk et al. (1998). Relevant studies conducted in developing countries are those by Lockheed and Zhao (1992), Jimenez and Sawada (1998) and King and Ozler (1998). Results show a mixed pattern, quite a few of these studies show no significant impacts of decentralization policies, but others do show outcomes that are in the expected direction, e.g. Woessmann (2001) and Mons (2004). The methodology that is employed may be a relevant factor in the heterogeneity of outcomes. The theoretical expectations behind the belief that

decentralization policies should have favourable impact on student performance are contested as well (see Mons 2005). The expectation, for example that decentralization is more efficient than centralization may be countered by experiences of new bureaucracy at lower administrative levels; and expected gains in excellence at the top of the student aptitude distribution may be countered by arguments of increased selectivity and inequality, as a result of market forces. Additional information on the effects of decentralization and school autonomy is provided in Chap. 9.

Evaluation and Accountability Arrangements

Freeing "process" and enforcing centralized control over outcomes is a pattern of decentralization that sets the scene for the development of accountability and evaluation arrangements in education. In this sense, the accountability movement could be subsumed under the larger issue of functional decentralization.

Expectations about the effectiveness of evaluation and accountability arrangements to enhance educational performance differ according to specific functional characteristics of evaluation and accountability arrangements.

A first major distinction could be made between evaluating for accountability purposes and evaluation for (organizational) learning and school improvement. Roughly, this distinction corresponds to evaluations being external, "summative" and judgmental in the case of accountability and evaluations being internal, "formative" and instrumental (i.e. not primarily intended to judge, but to diagnose and show ways to improve) in the case of organizational learning. In actual practice, mixed forms could exist as well, for example when national assessment information, collected on a census basis, is fed back to schools, as is for example the case in Italy.

In general terms, accountability refers to holding public institutions and services responsible for the quality and output of their performance. Glass (1972) states that accountability involves several loosely connected strands: "disclosure concerning the product or service being provided; product or performance testing; and redress for poor performance" (Glass 1972). The third element implies that accountability is not just a matter of providing and judging information but at least also 'foreshadows' actions by competent authorities in the sense of sanctions or rewards.

The first element—disclosure—requires that educational units, schools, in particular, provide information on their service provision, and make themselves "open" for external inspection and review. The second element distinguished by Glass stipulates that output and product information should be part of the disclosure on service provision and functioning. The third element emphasises that testing and review have implications in the sense of rewards and punishments for organisations. This relates accountability to incentive-based policies, like merit pay of teachers and output related financing of schools. Types of accountability are distinguished on the basis of who, or rather which kind of unit or stakeholder, is supposed to use the information that is disclosed by schools and teachers, and also who is supposed to apply the sanctions.

Elmore and Associates (1990) differentiate three "theories" of accountability on the basis of this question: who uses the information. They distinguish three types

- technical accountability, in which administrative units are supposed to take decisions on the basis of scientifically sound achievement measurements;
- the client perspective, in which the clients of education, like the parents of the pupils, "vote with their feet" in context of free choice of schools;
- the professional perspective; in which feedback on performance is basically used for professional development. "Accountability is, therefore, to be accomplished by deconstructing and reconstructing the meaning of schooling, collaborative planning, and co-operative teaching and learning" Elmore and Associates (1990), cited by MacPherson (1990, p. 7).

It could be maintained that only the two first forms can be seen as types of accountability. The "professional perspective" lacks the third element in Glass' basic definition, namely the application of rewards and sanctions. Moreover, what Elmore and Associates refer to as the professional perspective on accountability comes closer to the notion of "organizational learning" and the teacher as a reflective practitioner, as distinguished in the classical work of Argyris and Schön (1974). When specifying the professional perspective further, MacPherson also uses the term "empowerment" of teachers, which is more closely associated with school-based and school-initiated approaches to school improvement. In order to make the two "real" forms of accountability work, systems should have *evaluative capacity*, i.e. structural and technical facilities to realize the kinds of empirical disclosure and performance testing that accountability requires. With respect to the third defining element, the application of rewards and sanctions, which can be brought under the heading of incentive-based policies, research shows that there are often considerable limitations. When it comes to technical or administrative accountability, reviewers usually have to conclude that few examples of straightforward decision-making seem to exist. Cibulka and Derlin (1995), in their review of systems of school performance reporting (SPR), for example, say that "SPR is not considered very important by policy-makers or the general public". They conclude that it has not been demonstrated at all that "SPR can become a potent, effective policy-lever". Similar reservations have been based on empirical studies of the use that parents make of school performance information in choosing a school for their children (Bosker and Scheerens 1999). Nevertheless, there is evidence that accountability raises actual student achievement (e.g. Carnoy et al. 2003).

Next to accountability and organizational learning, certification could be distinguished as a third major functional area of educational evaluation. Certification is about formally regulating desired levels of quality of educational outcomes and provisions. Examinations, for example, at the end of lower secondary education, are there to certify students and to regulate what society can expect from those students (purposes of selection and stratification). At the same time, examination results could also be used in accountability contexts, for example, when pass-rates are used as performance indicators in judging the quality of schools. When the object of evaluation is not the individual student but the school as an organization, the term

accreditation is most commonly used, instead of certification. Quality management systems, like the well-known ISO norms can be applied to schools to check whether central work and managerial processes are in place and the organization is customer-oriented.

One could look upon evaluation and accountability arrangements at national level in terms of conditions that according to modern views on educational policy and management should simply be in place in order to open up educational service provision for external review and inspection, to formalize and standardize outcomes and as a characteristic of professional internal school management. The assumption that these provisions should also be seen as levers or drivers of educational effectiveness goes a step further. Summarizing the above, we have met the following hypothetical mechanisms according to which accountability and evaluation might also serve this latter, more ambitious demand

- administrative control in the sense of incentive-based policies, rewards, sanctions and public exposure of performance;
- consumer-based control, in which consumers who have free choice of schools, use performance information for school selection, and thus drive schools to compete in delivering good quality;
- organizational learning, in which good quality information guides internal improvement processes.

Although several studies show a positive association between evaluation, certification and accountability arrangements on the one hand and educational performance on the other, research outcomes do not unilaterally support any of the above underlying explanations (namely, the assumptions of administrative control, consumer-based control and the one about organizational learning). Research evidence does not quite support the first two explanations, mostly because pure forms are rarely implemented. Administrative control is frequently not applied strictly, and the basic conditions for performance-oriented client choice are rarely met. There is little hard evidence on the impact of school self-evaluation and organizational learning (Scheerens 2006). A more global explanation for the positive association between evaluation arrangements and performance might be that these provisions commonly enhance the result and outcome orientation in schools and thus stimulate achievement press (Bosker and Scheerens 1999). More details on empirical results concerning the effectiveness of evaluation and accountability policies will be provided in Chap. 9.

Integrated and Categorical Structures in Secondary Education

Recent analysis based on earlier waves of PISA have shown that comprehensive secondary school systems generally do better than categorical ones, both in terms of average performance levels, as in the sense of equity, Mons (2005), OECD (2005).

These results should be seen against the background of a long lasting debate about structural and institutional forms of differentiation of secondary schools, Husen (1979), Fend (1982), Tillmann (1989), Creemers and Scheerens (1988), Shavit and Blossfeld (1993), Crahay (2000), Zachary et al. (2002), Gorard et al. (2001), Duru-Bellat et al. (2004) and Mons (2004). The typology developed by Mons (2005) could be used to come to a more fine-grained distinction between systems that unify and differentiate to a different degree and on different facets.

The issue of integrated versus differentiated school structures, is related to the within-school phenomenon of streaming in the sense of ability grouping among classrooms at a specific grade level. Streaming might well take place within comprehensive secondary schools. Research results indicate that heterogeneous grouping is beneficial to lower achieving students. This might be explained by the fact that for lower achieving students, achievement standards will be kept high, when they are in classrooms where the average performance level is higher than their own (de Vos 1989). The lower performance of differentiated systems might also be due to the fact that "locking up" students with relatively low initial achievement in separate schools, leads to a too early fixed determination and lack of challenge. This effect is enhanced by lowering the age of first selection into a particular track of secondary schooling (Luyten et al. 2005).

Malleable Inputs and Processes at System Level

At the level of national or regional educational systems, it is somewhat arbitrary to categorize some policies in terms of ecology and others as directly malleable conditions. Accountability and evaluation policies, for example, would mostly require structural reform, as well as the initiation of specific procedures. To the degree that measures are more purely procedural within given structural arrangements (instead of being a mix of structural and procedural arrangements) and to the degree that measures, could, in principle, be limited to a one year policy program (such as, for example, the annual education budget), measures will be described as malleable inputs and processes rather than as ecological contextual conditions.

To the degree that educational systems decentralize important functions, the measures taken by the central level would be more oriented to setting general frames and restructuring the context rather than to taking direct control.

Three broad categories of educational policies, in the sense of malleable inputs and processes, within given structures will be considered:

- establishing the annual budget for education
- policy measures, in the sense of national programs, aimed at the improvement of educational outcomes and productivity
- equity-oriented policies
- teacher training

Annual Budget

Common indicators that show the annual investment in education are the percentage of the Gross Domestic Product (GDP) to be spent on education; and per student expenditure. Policy debates about increase in secondary costs, like administration costs, costs for support processes and "managerial overhead", would call for a decomposition of total expenditure on education in expenditure on primary processes of teaching and learning on the one hand and expenditure on secondary administration and support processes on the other.

Educational Programs Aimed at Enhancing Educational Quality

Countries are likely to have strongly varying educational policies, all ultimately aimed at enhancing the quality of education. Some of those would come under the heading of creating structural conditions and a quality enhancing context. No attempt will be made to try and find an overarching categorization scheme to make direct policy measures, internationally comparable. Instead, one dimension will be singled out; namely, the creation and implementation of national curricula that match national examinations and/or high-stakes testing programs.

Research outcomes such as Woessmann's (2001) as well as certain successes of centralized curricula for the core subject matter areas literacy and numeracy in the UK, would suggest that detailed, prescriptive national curricula are effective. This notion goes against the grain of an overall sentiment of decentralization in education, but can fit very well in a perspective of functional decentralization that was described earlier. Standard-based high stakes testing programs and standard-based examinations could have a similar effect. The discussion about detailed prescriptive national curricula resembles the discussion about the place of externally developed materials within the context of school improvement. Bottom up reform has dominated the fields of school improvement and curriculum implementation for decades, but was challenged by positive empirical evaluation outcomes of Comprehensive School Reform Programs. In the USA, a common core standard-based curriculum is currently an important issue (Schmidt et al. 2010). Ideas on curriculum alignment and opportunity to learn are becoming revitalized, in the wake of the movement towards common core standards (Polikoff and Porter 2014).

Equity-Oriented Policies

Equity in education can be described in terms of (in) equalities in educational outcomes, inputs and processes, across specific target groups. Between inputs and

processes on the one hand and outcomes a causal relationship may be hypothesized, in the sense that improvement of educational inputs and processes are expected to have impact on the equality of educational outcomes. The improvement in question could mean that inputs and outcomes are more equally distributed among target groups, or that they are specifically enhanced to better serve disadvantaged students (compensatory strategies). Conglomerates of input and process factors of schooling can be united into specific equity-oriented policies or programs. The core treatments of such programs are compensatory funding of schools in disadvantaged areas, and specifically developed curricula and lesson programs for disadvantaged learners. Extra money for disadvantaged schools is often used for class size reduction, or to appoint specially selected and trained teachers.

Initial Teacher Training

Initial teacher training is one of the basic conditions to maintain the well-functioning of educational systems. Yet, the research evidence, until very recently, does not support it as a system-level condition that does well in explaining achievement differences between countries.

Effects of teacher education, usually expressed in terms of formal qualifications, like having a BA or MA degree, or being certified to teach in a specific field, have traditionally been included in studies into "education production functions". In developed, industrialized countries, factors like formal qualifications do not appear to make much of a difference. In developing countries, such variables appear to be more often of significant impact. The explanation for this phenomenon is probably that the variation in formal teacher training in developed countries is usually quite limited, and teachers are more or less uniformly equipped to carry out their job. In developing countries, teacher preparation is less uniformly distributed. One could say that in developed countries, the impact of teacher education does not come out strongly from cross sectional and comparative studies, because of lack of variation in the independent variable. Formal characteristics of the way initial teacher training is organized in a country, for example, with respect to emphases on subject matter content and pedagogical content, and the degree to which teacher training institutions are formally monitored do not seem to seem to make much of a difference either (Scheerens 2014). More recently, there is attention for the finding that educational systems that show excellent performance, like Shanghai, Singapore and Finland, appear to be highly selective in the recruitment of teacher training candidates. Also, outcomes from the TEDS-M study (Teacher Education and Development Study in Mathematics) show important progress in establishing quality characteristics of initial teacher training programs, among others by demonstrating associations between opportunity to learn and teacher knowledge (Blömeke and Delaney 2012).

Conclusions About Policy Contexts and Actual Policies at System Level

At the level of national and regional educational systems, conditions that would ultimately make a difference with respect to student performance are again divided into three categories, antecedent conditions, context and ecology and direct manipulation of inputs and processes. At system level, the initial types of antecedents are constraints and facilities determined by the larger societal context. As such, economic, demographic, cultural and political conditions were distinguished. The second kind of contextual and "ecological" conditions were again described as a mixed set of given and manipulated conditions. At this level, structural reform measures concerning decentralization, accountability and evaluation and structural differentiation at secondary school level were distinguished as the most relevant categories. Finally, policies that are seen as more direct manipulations of inputs and processes were considered, and four kinds of direct policies were distinguished: funding formulas, quality improvement programs, equity-oriented policies and teacher training and recruitment.

Results are schematically summarized in Fig. 6.1 and Table 6.1.

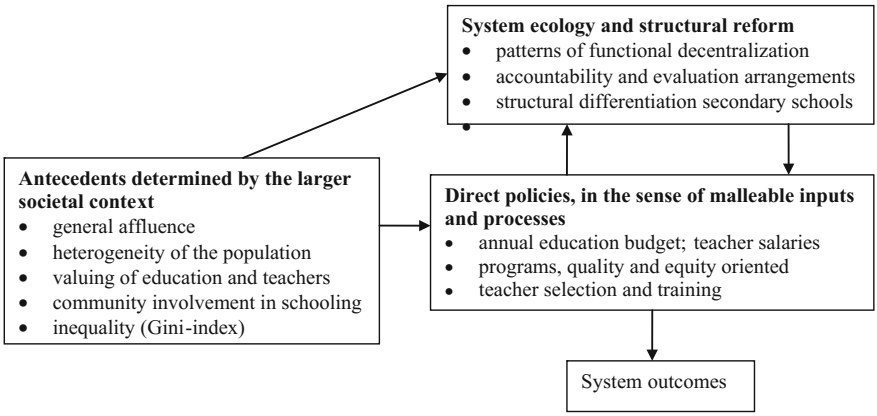

Fig. 6.1 System-level context and policies

Table 6.1 System-level policy context

Antecedents determined by the larger societal context	System ecology and structural reform	Direct educational policies in the sense of malleable inputs and processes
General affluence of the country/region GDP GDP per capita Annual unemployment quota (number of unemployed of the total potential working population) Percentage of people receiving social support of the potential adult working population **Heterogeneity of the population** Proportion of first, second and third generation immigrants Proportion of immigrants from developing countries **Status of teachers** Attractiveness of the teaching profession relative to other professions, as rated by the general public and teachers **Community involvement in schooling** Proportion of decisions that are taken by the local community/municipality (locus of decision-making survey) **Education relevant aspects of national cultures** **Societal (in)equality of country or region** (Gini index measures the extent to which the distribution of income (or consumption) among individuals or households within a country deviate from a perfectly equal distribution.)	**Functional decentralization** (de) centralization in curriculum, primary process of teaching, personal management and financial management **Evaluation and accountability arrangements** Variety of methods Institutional Infrastructure High or low stakes, nature of evaluation and accountability arrangements Whether or not a system has a standard-based examination at secondary level **Structural differentiation of secondary education** *Structural provisions* • Age of first selection of secondary education • Range of school types at secondary level (comprehensive versus categorical systems) • Percentage of schools that use streaming, in the sense of ability grouping per classroom	**Investment in education** Percentage of GDP spent on education Per pupil expenditure per ISCED level Teacher salaries **Teacher selection and training** Mandatory level of training Professional development Entrance levels teacher training programs **Quality-oriented reform programs** Common core standard-based curricula Evidence-based approaches (Comprehensive School Reform) Bottom-up reform **Equity-oriented policies** *Compensatory measures for disadvantaged learners* • Proportion of education budget spent on special programs • Priority of stratification of specific disadvantaged groups • Preferred strategies to compensate for educational inequality

References

Argyris, C., & Schön, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.

Bacharach, S. B., Bamberger, P., Conley, S. C., & Bauer, C. (1990). The dimensionality of decision participation in educational organizations: The value of a multi-domain evaluative approach. *Educational Administration Quarterly, 26*(2), 126–167.

Baumert, J., Carstensen, C. H., & Siegle, Th. (2005). Wirtschaftliche, soziale und kulturelle Lebensverhältnisse un regionale Disparitäten des Kompetenzerwerbs. In PISA-Konsortium Deustschland (Hrsg.) *PISA 2003. Der zweite Vergleich der Länder Deutschlands – Was wissen und können Jugendliche?* Münster/New York/Berlin/ München, 2005.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM—The International Journal on Mathematics Education, 44*, 223–247.

Bosker, R. J., & Scheerens, J. (1999). Openbare prestatiegegevens van scholen; nuttigheid en validiteit. *Pedagogische Studiën, 76*(1), 61–73.

Bray, M. (1994). Centralization/decentralization and privatization/ publicization: Conceptual issues and the need for more research. In W. K. Cummings & A. Riddell (Eds.), *Alternative policies for the finance, control, and delivery of basic education* (pp. 817–824). Special issue of the *International Journal of Educational Research, 21*(8).

Bryk, T., Thum, Y., Easton, J., & Luppescu, S. (1998). *Charting Chicago school reform*. Boulder: Westview Press.

Caldwell, B. J., & Spinks, J. M. (1988). *The Self-managing School*. London: Falmer Press.

Carnoy, M., Elmore, R., & Siskin, L. (2003) (Eds.). *The new accountability. high schools and high-stakes testing*. New York/London: Routledge Falmer.

Cibulka, J. G., & Derlin, R. L. (1995). State educational performance reporting policies in the U. S.: Accountability's many faces. *International Journal of Educational Research, 23*(6), 479–492.

Crahay, M. (2000). *L'école peut-elle être juste et efficace? De l'égalité des chances à l'égalité des acquis*. Bruxelles: Editions De Boeck Université.

Creemers, B. P. M., & Scheerens, J. (1988). Political and organizational influences on middle school evaluations. *International Journal of Educational Reserch, 12*(5), 561–569.

de Vos, H. (1989). A rational-choice explanation of composition effects in educational research. *Rationality and Society, 1*(2), 220–239.

Duru-Bellat, M., Mons, N., & Suchaut, B. (2004). *Caractéristiques des systèmes éducatifs et compétences des jeunes de 15 ans : l'éclairage des comparaisons entre pays.* Rapport pour le M.E.N., DIJON: IREDU-CNRS.

Elmore, R. F. and Associates. (1990). *Restructuring school: The next generation of educational reform.* Jossey Bass: San Francisco.

Fend, H. (1982). *Gesamtschule im Vergleich: Bilanz der Ergebnisse des Gesamtschulversuchs*. Weinheim: Beltz.

Fuchs, T., & Woessmann, L. (2004). *Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school*. Munich: CESIFO Working University of Munich.

Fullan, M., & Watson, N. (2000). School-based-management: Reconceptualizing to improve learning outcomes. *School Effectiveness and School Improvement, 11*(4), 453–473.

Glass, G. V. (1972). The many faces of educational accountability. *Phi Delta Kappan, 53*, 636–639.

Gorard, S., Fitz, J., & Taylor, C. (2001). School choice impacts: What do we know? *Educational Researcher, 30*(7), 18–23.

Husen, T. (1979). *L'école en question*. Bruxelles: Mardaga.

James, E. (1994). The public-private division of responsibility in education. In W. K. Cummings & A. Riddell (Eds.), Alternative policies for the finance, control, and delivery of basic education. Special issue of the *International Journal of Educational Research*, *21*(8), 777–783.

Jimenez, E., & Sawada, Y. (1998). *Do community-managed schools work? An evaluation of El Salvador's EDUCO program.* Working Papers Series on Impact Evaluation of Education Reforms No. 8. Washington, DC: World Bank.

King, E. M., & Ozler, B. (1998). *What's Decentralization got to do with learning? The case of Nicaragua's School Autonomy Reform*. Paper Presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 13–17, 1998.

Leithwood, K., & Menzies, T. (1998a). Forms and effects of school-based management: A review. *Educational Policy, 12*(3), 325–346.

Leithwood, K., & Menzies, T. (1998b). A review of research concerning the implementation of site-based management. *School Effectiveness and School Improvement, 9*(3), 233–285.

Lockheed, M., & Zhao, Q. (1992). *The empty opportunity: Local control of secondary schools and student achievement in the Philippines*. Working Paper WPS 825. Washington DC: World Bank.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B. U., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD.

MacPherson, R. J. S. (1990). *Educative accountability: Theory, practice, policy and research in educational administration*. Oxford: Pergamon.

Malen, B., Ogawa, R.T., & Kranz, J. (1990). What do we know about school-based management?: A case study of the literature— A call for research. In W. H. Clune & J. F. Witte (Eds.), *Choice and control in American Education (Vol. 2): The practice of choice, decentralization and school restructuring* (pp. 289–342). London: Falmer Press.

Maslowski, R., Scheerens, J., & Luyten, H. (2007). The effect of school autonomy and school internal decentralization on students' reading literacy. *School Effectiveness and School Improvement, 18*(3), 303–334.

Mons, N. (2004). *De l'école unifiée aux écoles plurielles*: *évaluation internationale des politiques de différenciation et de diversification de l'offre éducative*. Doctorat en sciences de l'éducation, IREDU-Université de Bourgogne.

Mons, N. (2005). Doit-on former ou sélectionner les élites? Une comparaison internationale des politiques éducatives. *Revue Internationale d'Education de Sèvres*, 39.

Murphy, J., & Beck, L. G. (1995). *School-based management as school reform: Taking stock*. Newbury Park: Corwin Press.

OECD. (2005). *School Factors Related to Quality and Equity: Results from PISA 2000*. Paris: OECD. Available from: http://www.pisa.oecd.org/dataoecd/15/20/34668095.pdf.

OECD. (2010). *The high cost of low performance: The long-run economic impact of improving PISA outcomes*. Paris: OECD Publishing.

Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*, *36*, 399–416.

Rideout, W. M., & Ural, P. (1993). *Centralized and decentralized models of education: Comparative studies*. Centre for Policy Analysis: Development Bank of Southern Africa.

Scheerens, J. (2006). New system level indicators. Discussion paper for the Network C meeting in Oslo 29–30 May 2006.

Scheerens, J. (2014). Teacher training and professional development in Europe: In search of effects on educational performance. *Key note address at the German Rector's conference*, Essen, 21 January, 2014.

Schmidt, W. H., Cogan, L. S., & McKnight, C. C. (2010). Equality of opportunity: Myth or reality in U.S. Schooling. *American Educator, Winter, 2010–2011*, 12–19.

Shavit, Y., & Blossfeld, H. -P. (1993). *Persistent inequality*. London: Westview Press.

Summers, A. A., & Johnson, A. W. (1996). The effects of school-based management plans. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's Schools: The role of incentives* (pp. 84–91). Washington, DC: National Academy Press.

Tillmann, K. J. (1989). *Sozialisationstheorien: Eine Einführung in den Zusammenhang von Gesellschaft, Institution und Subjectwerdung*. Reinbek bei Hamburg: Rowohlt.

van Amelsvoort, H. W. C. H., & Scheerens, J. (1997). Policy issues surrounding processes of centralization and decentralization in European education systems. *Educational Research and Evaluation, 3*(4), 340–363.

Wang, M. C., & Walberg, H. J. (2001). *Tomorrow's Teachers*. CA: McCutchan Publishing Company.

Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education: The school, the state and the market*. Buckingham: Open University Press.

Winkler, D. R. (1989). *Decentralisation in education: An economic perspective*. Washington: The World Bank, Population and Human Resource Development.

Winkler, D. R., & Gershberg, A. I. (2000). *Education decentralization in Latin America: The effects on the quality of schooling* (LCSHD Paper Series No. 59). Washington, DC: World Bank.

Woessmann, L. (2001). *Schooling resources, educational institutions, and student performance: The international evidence*. (Kiel Working paper No. 983). Kiel, Germany: Kiel Institute of World Economics.

World Bank. (2003). *World Development Indicators 2003*. CD-ROM. Washington, DC.

Zachary, M. D., Dupriez, V., & Vandenberghe, V. (2002). Comparaison internationale de l'efficacité et de l'équité des systèmes d'enseignement. In D. de la Croix, F. Docquier, C. Mainguet, S. Perelman & E. Wasner (Eds.), *Capital humain et dualisme sur le marché du travail* (pp. 259–279). Bruxelles: De Boeck.

Part II
Quantitative Research Results

Introduction to Part II

In Part II of this book results of quantitative research syntheses and meta-analyses are reviewed. The central chapter (Chap. 8) describes meta-analyses of school and instructional effectiveness carried out as part of the research programme on educational effectiveness at the University of Twente (Scheerens et al. 2007, 2013). In these meta-analyses the variables that were identified in Part I, as summarized in Tables 3.6 and 5.6, were used as a basis for selecting empirical studies. In this way the results provide information on the quantitative effects of the variables that resulted from the qualitative review in Part II. Before the meta-analyses are presented, Chap. 7 provides a brief introduction to the kind of measures of effect sizes that are used in educational effectiveness research. Chapter 9 provides miscellaneous results on system effectiveness and in Chap. 10 results of our "own" meta-analyses are compared to other "multi-variable" educational effectiveness meta-analyses and to more recent "single-factor" meta-analyses carried out by ourselves and others. This chapter concludes with an attempt to make up the balance of these "glimpses of the educational effectiveness knowledge base", an effort that will be continued in Part III, in which the research evidence is addressed from the perspectives of theory and educational policy and practice.

References

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organization and Management.

Scheerens, J., Witziers, B., & Steen, R. (2013). A Meta-analysis of School Effectiveness Studies. *Revista de Educacion*, *2013*(361).

Chapter 7
Effects and Effect Sizes in Educational Effectiveness Research

Keywords Effect sizes · Variance components · Absolute educational effects · Value-added methods · Growth curves · Regression discontinuity design · Path analysis and multi-level structural equation modelling · *d*-coefficient · Composition effects

Introduction

In educational effectiveness research we are not just concerned with effect measures, like average achievement scores, or certain attainment standards, but always also with a condition or "cause" to which variation in effects can be attributed. Roughly there are two main questions. The first question is what difference it makes to go to one school or the next, to be part of national education system A or B, or which classroom a student attends. Such effects are usually indicated as school effects, classroom effects, or country effects and depend on the organizational tier or unit a student belongs to. Apart from the three units (schools, classes, countries) other units can be used in this way, e.g. principals, teachers, grade levels, municipalities, districts and regions. The second question is about identifying variables that further characterize inputs and processes within units, and assessing their association with effect measures. Other relevant distinction with respect to effectiveness measures are whether they are relative or absolute, and whether they are presented as "gross" or adjusted ("net") measures.

The "Effect" of Belonging to a Unit at a Certain Level: Decomposition of Variance

It is quite common to express school effects in terms of the percentage of total between student variance in achievement "explained" or tied by the school. Scheerens and Bosker (1997, 79) estimate the "gross" school effect, defined in this

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_7

way at 9 %. This effect may differ according to the grade level and the subject matter area that is assessed. Usually effects are larger for mathematics and science than for reading literacy. It should be noted that school effects expressed as percentages of variance explained are relative measures of educational quality. As a matter of fact they might as well be interpreted as a manifestation of equity in the sense of the dispersion of school average achievement in a country. Effect sizes depend on the variability or homogeneity of schools in a particular context, e.g. a particular country. School effects in terms of percentage of total variance tied by the school can also be expressed in more conventional effect size coefficient, such as Cohen's *d* (Tymms 2004). The gross school effect of 9 % explained variance, confirms to an effect size estimate of 0.30 (Scheerens and Bosker 1997, ibid., p. 79).

In other applications, "percentage of variance explained" indicators are used to assess the relative contribution of schools versus other units, classrooms in particular. When such comparisons are made it should be noted that the results reflect variability rather than absolute measures of quality. When the variance between classes appears to be higher than the variance between schools, this means that, in the particular setting, it makes more of a difference in which classroom a student is located, than to which school this student goes. The reference to a "particular setting" might well be the country-level context. This is illustrated in Table 7.1, cited from Scheerens et al. (1989).

Table 7.1 Estimates of the variance explained by schools and classes, cited from Scheerens et al. (1989), based on the second international mathematics study by I.E.A.

Private country	Classroom variance component	School variance component
15 Belgium (Flemish)		0.50
16 Belgium (French)		0.64
22 Canada (British Columbia)		0.27
25 Canada (Ontario)	0.18	0.09
39 Finland	0.45	0.002
40 France	0.17	0.06
43 Hong Kong		0.51
44 Hungary		0.30
50 Israel	0.22	0.10
54 Japan		0.08
59 Luxembourg	0.29	0.15
62 The Netherlands		0.67
63 New Zealand	0.45	0.01
72 Scotland	0.34	0.12
76 Sweden	0.45	0.00
79 Thailand		0.39
81 USA	0.46	0.10

Note Estimation of the variances expressed in terms of the intra-class correlation coefficient, for all countries, assuming schools are sampled at random within countries and classrooms are sampled at random within schools

Some countries, Sweden and Finland, show a high between classrooms component and a very low between schools component, which means that schools in these countries are very homogeneous, but that there is more variation between schools. In the USA data from this study there was a strong between classrooms component and still a sizeable between schools component. Countries like Belgium and the Netherlands have high between school variance due to a categorical secondary school system.

Hattie (2009, p. 18) provides the following comparative estimates of the difference made by the student, the teacher and the school, in terms of the effect sizes (*d*-coefficient): student 0.44, teacher, 0.49 and school 0.23. He concludes that teachers/classrooms matter more than schools. This pattern is usually supported in other studies; Kyriakes and Luyten (2009) report percentages of variance explained of 69.2 % for students, 17.7 % for classrooms and 13.1 % for schools. Opdenakker and Van Damme (2000) found that 12.8 % of the variance was between schools, 18.07 % between teachers, 14.71 % between classes and 54.42 % between students. These authors point to the fact that omitting intermediary levels (classes, teachers) in multi-level analyses tend to lead to an overestimating of the effect of the next above level, in this case schools.

International studies allow adding a variance component (or an effect size) for the country level.

Luyten et al. (2005) presented the following decomposition of total between student variations, based on PISA 2000 data (Table 7.2).

The total number of OECD countries in PISA 2000 was 27. The overall pattern the table shows is that a sizeable amount of variation lies between countries, although less than between schools.

In a study in which 12 educational systems from the UK and Australia were compared on math achievement progress, Tymms and Wildy (2014) found a higher percent of variance tied by the system level (13 %) than by the class level (7 %). However, intermediary level components, the classroom and the year group, taken together tied more variance (20 %) than did the system level.

Based on the data from TIMSS, Kyriakides (2006) found that 20 % of the variance was associated with countries.

Table 7.2 Percentage of variance in student performance in reading, mathematical and scientific literacy in OECD countries

	Percentage at country level	Percentage at school level	Percentage at student level
Reading	8	15	57
Mathematics	16	31	54
Science	10	32	59

Results from PISA 2000, cited from Luyten et al. (2005, p. 116)

Approximation of Absolute School Effects

As noted above, the methods to establish "school effects" discussed so far relate only to relative differences between schools. As a matter of fact the effect of schooling is not measured, but rather the variation between schools (Luyten 2007, 173).

A first important step to get a more satisfactory interpretation of school effects is the use of gain scores, or growth curves of student progress as effect variables. Such assessments are usually referred to as "value-added measures". Applying these is a step forward in getting a more convincing interpretation on school effects, namely by trying to separate "what schools do" from other external influences. This is most relevant for educational effectiveness research, as will be pointed out in the next section, but also important for practical applications, such as judging school performance, as the rankings come closer to expressing the merit, rather than mixtures of merit and social intake characteristics.

In order to further rule out that effects of schooling, even when they are computed as sophisticated value-added measures, still reflect external influences, more powerful counterfactuals need to be created. The challenge is to find conditions of "non-schooling" as a basis for assessing the effect of schooling. This is attempted, for example, when evaluating the effects of "summer learning" and by specific applications of regression discontinuity designs. Citing Luyten (2007):

"The first measure (the effect of summer learning, J.S.) is based on a comparison of the rate of learning during the school year and the rate of learning during the summer holidays". Downey et al. (2004) proposed seasonal "impact" as a measure of school effectiveness. This is the difference between the school year and the summer gain rates. The effect of schooling is supposed to be zero if the learning rate during the school year equals the rate during the summer vacation. The second measure capitalizes on the fact that in most countries school admission is primarily determined by the date of birth. In this case, the between grade differences are treated from a selection bias perspective. Differences between grades are conceived as the result of different treatments, but the approach takes into account that some students may have been assigned to a lower or higher grade than expected given their age, reasons that also affect their scores on the outcome variables (like learning or behaviour problems). If the criteria that determines who receives a "treatment" (upper grade) and who is placed in the control group (lower grade) can be exactly specified the regression-discontinuity approach which can be applied to assess the effect of the "treatment". By taking the impact of these criteria into account in the data analysis, we can assess the impact of the treatment most accurately (Shadish et al. 2002; Trochim 1984). If the cut-off date that determines what grades students should be placed in is strictly followed, the effect of one-year schooling can be assessed by adjusting the difference in achievement between two (or more) grades in a row for the effect of age. This adjusted difference reflects the gap between the oldest students in the lower grade and the youngest in the upper grade (Cahan and Davis 1987; Luyten 2006). In this case the effect of schooling is supposed to be zero if the difference in achievement between grades is completely accounted for by

the age of the students. The main practical complication of this approach is that it assumes a strict adherence to the cut-off date. In many educational systems this is not the case. Grade repeating, especially, is a common phenomenon in a wide range of countries" (Luyten 2007, 173/174).

Applying these methods yielded effect sizes (Cohen's *d*) for one year of schooling, in the order of 0.99 (reading speed), 0.84 (mathematics) and 0.33, when the approach of summer learning was applied (Luyten 2007). In another application of the regression discontinuity design Kyriakides and Luyten (2009) found effect sizes of 0.87 for both mathematics and reading.

When comparing these values with the effects sizes that were found, when applying the variance decomposition approach, they are definitely much higher (a difference between high and low/moderate, when applying Cohen's standards).

Opening the Black Boxes of Unit Effects

How much differences schools, classes or even country-level educational systems make relatively and in a more absolute sense was discussed in the previous section. The next question is to what extent these effects of belonging to a certain unit, or of being treated or not by schooling, can be "explained" by more specific treatments of malleable factors that are active at specific unit levels. This is more at the core of what educational effectiveness research is all about. As a way of introducing the issue, the following citation from Hanushek and Rivkin, about teacher effects, is quite striking:

"Literally hundreds of research studies have focused on the importance of teachers for student achievement. Two key findings emerge. First, teachers are very important. No other measured aspect of schools is nearly as important in determining student achievement. Second, it has not been possible to identify any specific characteristics of teachers that are reliably related to student outcome". (Hanushek and Rivkin 2010). In the same vein Rivkin et al. (2005) conclude that "teachers have powerful effects on reading and mathematics achievement, though little of the variance in teacher quality is explained by observable variables, such as education or experience".

The distinction is clear: although we have ample evidence that "teachers matter", in the sense if being taught by one teacher or the next, our knowledge on how they matter, may be limited.

When we move from assessing overall unit effects to establishing more specific explanations of these the following conceptual steps are to be taken:

- identification of malleable, policy amenable variables that are assumed to be associated with educational outcomes;
- separating the influences of malleable variables from "given" antecedent conditions, like student background characteristics.

The identification of malleable variables that are associated with educational outcomes has been treated in Part 1 of this book. It has resulted in a set of school and classroom levels variables, (summarized in Tables 3.6 and 5.6), which will be used in meta-analyses that will be described in the next chapter.

Separating the influences of malleable and background variables has been touched upon in the previous section, when "value added" outcome measures, in other words effect measures adjusted on the basis of relevant covariables, was discussed. Whereas in practical applications sometimes a case can be made to use "gross" measures of school effects,[1] instead of net effects, in educational effectiveness *research* adjusting for relevant background conditions is a must.

In standard multi-level analysis, a first model provides estimates of variance components of the level units, next, student-level background characteristics are fitted, followed by malleable variables defined at class and/or school level. The regression coefficients for the different categories of malleable variables reflect the association of the particular variable, net of background conditions and other independent variables used in the analysis. Such regression coefficients can be expressed as correlations or effect sizes (like Cohen's *d*) for each of the independent variables of interest. Interaction effects at one particular level, or across levels, can be included as well. This is the most frequently applied methodology to assess the effectiveness of specific malleable school and classroom-level variables to date, and the large majority of the studies used in the meta-analyses described in the next chapter used this methodology. What this methodology enables is a more extended decomposition of variance components, including unexplained variance and school composition effects. Figure 7.1 based on a study by Brandsma (1993) illustrates this.

This study was conducted in Dutch primary schools, and used a pre-test as student background variable. The 10 % "net" between school variance is the percentage of variables that was tied by all the substantive effectiveness enhancing variables in the model. Note also the large effect of the covariables at individual student level (50 %) and at school level (16 %).

Some innovations to this standard methodology are discussed in the subsequent points.

Growth Curves

Using gain scores or several measuring points on the effect variable can be seen as a special case of adjustment in the procedure described above. Seen from this perspective using pre-test information can be seen as choosing a powerful covariable. Empirical evidence seems to indicate that applying growth curves, in which

[1]For example, when school choice by parents is at stake, as compared to school Inspectorates assessing the merit of a school.

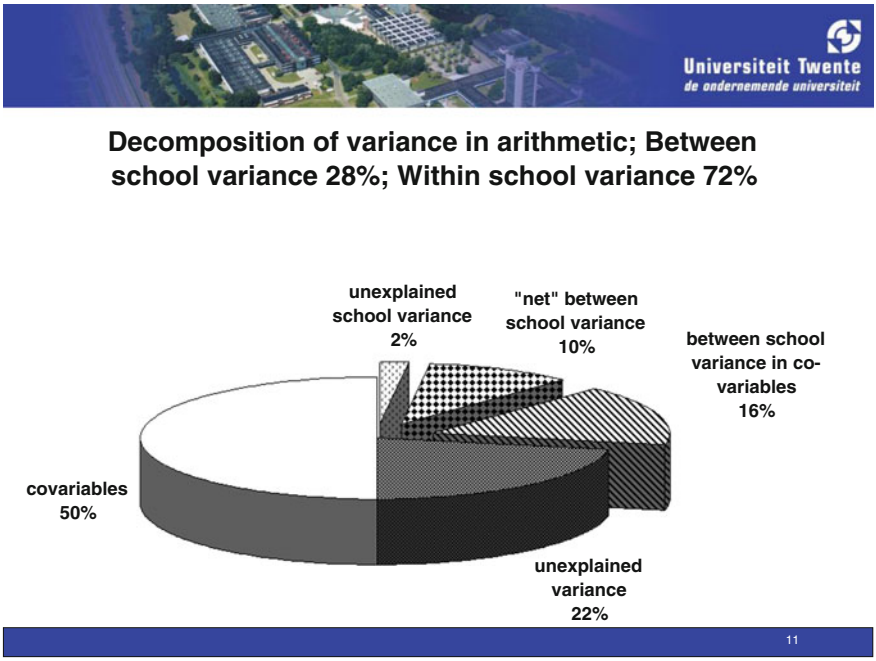

Fig. 7.1 Decomposition of variance in arithmetic. From Brandsma (1993)

progress is measured over more than two points in time, produces higher effect sizes as compared to pre-test–post-test or other covariable adjustments. Guldemond and Bosker (2009, 266) report school effects amounting to 30–50 % of the variance in growth curves, while Rowan et al. (2002) mention *d*-type effects of 0.72–0.85 for overall teacher effects, with more specific teacher and teaching characteristics producing *d*-type effects of about 0.10.

Analyzing Absolute Effects

Another approach to enhance the effect estimation of malleable school and teaching conditions might consist of determining "absolute" effects of schooling, as described in the previous section, and specifying these for each school. The variance of such "absolute" school scores could then be further decomposed by means of multi-level analyses. This does not seem to be common practice, as yet, however. An application is provided in Heck and Moriyama (2010), in a study that assessed school leadership effects.

Application of Path Analysis and Structural Equation Modelling

In the standard approach, effects of malleable variables are assessed "one by one", categorized by level. Following the structure of the basic conceptual model, presented in Chap. 1, the influence of certain variables may be indirect and mediated by third variables (a mediator, in addition to the independent variable of interest and the effect variable). Path analysis and structural equation modelling allow for the computation of indirect effects, involving one or more mediating variables. Most applications are currently found in studies of school leadership effects (Scheerens 2012; Hendriks and Scheerens 2013). In principle, this methodology has the potential of bringing the conceptual models of educational effectiveness closer to the operational models that are used in data analysis. The more so in the case of multi-level structural equation modelling (Scheerens et al. 2014).

Program Evaluations and (Quasi) Experiments to Assess Educational Effectiveness

An alternative to assess the effects of malleable school and teaching variables on the basis of survey research methods and correlational types of analyses, as the standard approach described in the above, is the application of experiments, either "true" randomized field trials or quasi-experiments. This approach is more easily realizable for classroom-level studies of teaching approaches than for school-level variables. Seidel and Shavelson (2007) found evidence showing that quasi-experimental studies of teaching approaches show higher effect sizes than correlational studies. For school-level variables the experimental condition is likely to get the scope of a program. Experimental evaluations of Comprehensive School Reform Programs in the US provide interesting examples. Here the treatment consists of a combination of several school and teaching variables that have individually been shown to be associated with higher achievement, like instructional leadership, cooperative learning, frequent monitoring of student progress and parental involvement. Evaluations and meta-evaluations of such programs tend to show relatively small, but consistent effects (Borman et al. 2003, 2007).

References

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125–230.

Borman, G., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of success for all. *American Educational Research Journal, 44*(3), 701–731.

Brandsma, H. P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs (Characteristics of primary schools and the quality of education)*. Groningen: RION.

Cahan, S., & Davis, D. (1987). A between-grades-level approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal, 24*(1), 1–12.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*, 613–635.

Guldemond, H., & Bosker, R. J. (2009). School effects on students' progress—adynamic perspective. *School Effectiveness and School Improvement, 20*(2), 255–268.

Hanushek, E., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review: Papers and Proceedings, 100*, 267–271.

Hattie, J. (2009). *Visible learning*. Abingdon: Routledge.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: a regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Hendriks, M. A., & Scheerens, J. (2013). School leadership effects revisited: A review of empirical studies guided by indirect-effect models. *School Leadership and Management, 33*(4), 373–394.

Kyriakides, L. (2006). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis for TIMSS 199 data. *Educational Research and Evaluation, 12*(6), 513–534.

Kyriakides, L., & Luyten, H. (2009). The contribution of schooling to the cognitive development of secondary education students in Cyprus: an application of regression discontinuity with multiple cut-off points. *School Effectiveness and School Improvement, 20*(2), 167–186.

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. *Oxford Review of Education, 32*(3), 397–429.

Luyten, H. (2007). Assessing the effect of schooling, alternatives for the focus on school differences. In J. Scheerens, H. Luyten, R. Steen, & Y. Luyten-de Thouars (Eds.), *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organization and Management.

Luyten, H., Scheerens, J., Visscher, A., Maslowski, R., Witziers, B., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000.* Paris: OECD.

Opdenakker, M.-C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11*(2), 165–196.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teacher, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rowan, B., Correnti, R., & Miller, J. (2002). What large-scale survey research tells us about teachers effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*(8), 1525–1567.

Scheerens, J. (Ed.). (2012). *School leadership effects revisited. review and meta-analysis of empirical studies.* Dordrecht: Springer.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., Luyten, H., Glas, C. A., Jehangir, K., & Van den Bergh, M. (2014). System level indicators. Analyses based on PISA 2009 data. Internal report. Enschede: University of Twente.

Scheerens, J., Vermeulen, C. J. A. J., & Pelgrum, W. J. (1989). Generalizibility of instructional and school effectiveness indicators across nations. *International Journal of Educational Research, 13*(7), 789–799.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Trochim, W. M. K. (1984). *Research design for program evaluation, the regression-discontinuity approach*. London: Sage Publications.

Tymms, P. (2004). Effect sizes in multi-level models. In I. Schagen & K. Elliot (Eds.), *The use of effect sizes in educational research*. National Foundation for Educational Research: Slough.

Tymms, P., & Wildy, H. (2014). The progress of pupils in their first school year across classes and educational systems. *British Educational Research Journal.* (online) doi:10.1002/berj.3156.

Chapter 8
Meta-Analyses of School and Instructional Effectiveness

Keywords School effectiveness · Instructional effectiveness · Achievement · Orientation · High expectations · Educational leadership · Consensus and cohesion among staff · Curriculum quality/opportunity to learn · School climate · Evaluative potential · Parental involvement · Evaluative potential · Classroom climate · Effective learning time · Achievement orientation · Educational leadership · Differentiation · Classroom organization · Learning environment · Clear and structured teaching · Activating teaching · Teaching learning strategies · Challenging teaching · Supportive teaching · Feedback · Evaluation at classroom level · Teacher characteristics · Adaptive teaching · Specific teaching practices · Use of teaching materials · Integrative teaching approaches

Introduction

In this chapter, quantitative analyses are described that were conducted on the basis of a data set that combined studies from two previous meta-analyses, namely those by Scheerens and Bosker (1997) and by Scheerens et al. (2005). Both data sets were rechecked and scrutinized, which in some cases led to a different "scoring" of the basic effect sizes (for a further explanation on this checking process see Annex 1 to this chapter). Only associations of school and instructional variables with cognitive educational achievement were used in the analyses.

The sizes of the database for the school-level factors and instructional-level factors are shown in Tables 8.1 and 8.2, respectively.

Statistical analyses for this chapter were carried out by Rien Steen.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_8

Table 8.1 Numbers of studies and replications for the school-level variables

	Publications	Replications
"Old" studies 1985–1994	83	511
"New" studies 1995–2005	72	700
Total	155	1211

"Replications" are associations between the school variable in question and student achievement

Table 8.2 Numbers of studies and replications for the instructional variables

	Publications	Replications
"Old" studies 1985–1994	44	126
"New" studies 1995–2005	130	1786
Total	177	1912

"Replications" are associations between the school variable in question and student achievement

The Variables Used in the Meta-Analyses

The school and instructional variables that were used in the second round of meta-analyses provide a reasonable match with the key effectiveness-enhancing variables that were identified in Chaps. 3 and 5. The variables that formed the basis for the meta-analyses are summarized in Tables 8.3 and 8.4.

As described in Chap. 5, an alternative categorization for the teaching variables, in terms of six more global factors were proposed. This categorization is shown in Table 8.5.

Table 8.3 Components of 13 effectiveness-enhancing school factors

Factors	Components
Achievement, orientation, high expectations	• Clear focus on the mastering of basic subjects • High expectations (school level) • High expectations (teacher level) • Records on pupils' achievement
Educational leadership	• General leadership skills • School leader as information provider • Orchestrator of participative decision-making • School leader as coordinator • Meta-controller of classroom processes • Time educational/administrative leadership • Counsellor and quality controller of classroom teachers • Initiator and facilitator of staff professionalization
Consensus and cohesion among staff	• Types and frequency of meetings and consultations • Contents of cooperation • Satisfaction about cooperation • Importance attributed to cooperation • Indicators of successful cooperation

(continued)

Table 8.3 (continued)

Factors	Components
Curriculum quality/opportunity to learn	• The way curricular priorities are set • Choice of methods and text books • Application of methods and text books • Opportunity to learn • Satisfaction with the curriculum
School climate	(a) *Orderly atmospheres* • The importance given to an orderly climate • Rules and regulations • Punishment and rewarding • Absenteeism and drop out • Good conduct and behaviour of pupils • Satisfaction with orderly school climate
	(b) *Climate in terms of effectiveness orientation and good internal relationships* • Priorities in an effectiveness-enhancing school climate • Perceptions on effectiveness-enhancing conditions • Relationships between pupils • Relationships between teacher and pupils • Relationships between staff • Relationships: the role of the head teacher • Engagement of pupils • Appraisal of roles and tasks • Job appraisal in terms of facilities, conditions of labour • Task load and general satisfaction • Facilities and building
Evaluative potential	• Evaluation emphasis • Monitoring pupils' progress • Use of pupil monitoring systems • School process evaluation • Use of evaluation results • Keeping records on pupils' performance • Satisfaction with evaluation activities
Parental involvement	• Emphasis on parental involvement in school policy • Contacts with parents • Satisfaction with parental involvement
Classroom climate	• Relationships within the classroom • Order • Work attitude • Satisfaction
Effective learning time	• Importance of effective learning • Time • Monitoring of absenteeism • Time at school • Time at classroom level • Classroom management • Homework
Structured instruction	• Importance of structured instruction • Structure of lessons

(continued)

Table 8.3 (continued)

Factors	Components
	• Preparation of lessons • Direct instruction • Monitoring
Independent learning	No sub-components
Differentiation	• General orientation • Special attention for pupils at risk
Reinforcement and feedback	No sub-components

Table 8.4 Teaching variables

Instruction characteristics	Indicators
(1) Learning time	Time on task, effective use of teaching time, homework, mastery learning
(2) Opportunity to learn	Content covered, opportunity to learn
(3) Classroom organization	Classroom management, discipline, control
(4) Orderly, functional learning environment	Learning climate, classroom climate, achievement pressure[a], Mastery orientation, performance orientation[a]
(5) Clear and structured teaching	Direct teaching, structured teaching, teacher demonstrations, teaching for basic skills, clarity
(6) Activating	Cooperative learning, situated learning, discovery learning, peer tutoring, student experiments, hands-on activities, group work, individual work, individual learning, discussions
(7) Learning to use learning strategies	Cooperative learning strategy training, problem-solving, metacognitive training, scientific inquiry training, thinking aloud training, concept mapping, organizing/structuring methods, language acquisition training, phonemic awareness training, reading strategies, writing strategies, formal learning strategy training
(8) Challenge	Cognitive activation, orientation towards understanding, active student engagement, authentic contexts, relevance to students, language level, varying representation formats
(9) Support (mutual respect)	Quality of teacher–student interactions, student–student interactions, teacher support
(10) Feedback/monitoring	Feedback, monitoring, individual frame of reference
(11) Evaluation of goals/attainment	Assessments, tests
(12) Adaptive teaching	Variable teaching methods, adaptive teaching, orientation towards individual learning processes, choice, taking into account student prerequisites
(13) Practice	Drill, repetitions, applications
(14) Material	Quality of curriculum, textbooks, use of computers
(15) "Integrated" instructional concepts	Constructivist instruction, inductive teaching, concept-oriented/integrated instruction

[a]Reverse coded for summary

Table 8.5 A six-way categorization of teaching variables

Global teaching factor	Instructional characteristics
(1) Curricular	Time, opportunity to learn, textbooks, media
(2) Teacher-orchestrated classroom management	Classroom organization, orderly and functional learning environment
(3) Structured teaching strategy	Clear and structured teaching. performance orientation, drill and repetition, application, mastery learning orientation
(4) Constructivist-oriented teaching strategy	Activating, learning to use learning strategies, challenge, integrated instructional concepts
(5) Climate, support, positive interactions	Classroom climate, positive interactions
(6) Evaluation and monitoring	Feedback, assessment, tests

Methods

Literature Search

A meta-analysis relies on collecting as many studies as possible regarding the topic of interest. The search methods included searches on the Web of Science, and the ERIC and ERA databases. The search was focused at articles published between 1985 and 2005. In addition, the literature database of ECER conferences was examined. In the search the following key words were used: school effectiveness, learning results, effectiveness, effective teaching, effective instruction, teacher effectiveness, educational effectiveness, school effectiveness and student achievement. Finally, recent reviews and books on school effectiveness were checked in order to find additional relevant literature ("snowball method").

The first step of this search resulted in several hundreds of publications. From these publications, about one-third appeared not to be useful for our purposes, while from one-sixth of all publications it could not be determined whether or not they contained useful information. These were articles that appeared to be inaccessible. This left us with 155 articles that contained information relevant for the purposes of our study in the domain of school effectiveness and 177 articles in the domain of teaching effectiveness. These articles were analysed with regard to effect size presented on student achievement outcomes and relevant school and teaching effectiveness variables, while at the same time data were collected on particular study characteristics.

Types of Meta-Analysis

A meta-analysis can be conducted in different ways. In this study two types of meta-analyses were carried out. First, a so-called "vote-counting" procedure took

place was applied to the school-level variables. Vote counting comes down to counting the number of significant (positive and negative) associations between a dependent variable and a specific independent variable of interest from a given set of studies. More specifically, it was examined for each selected study whether or not the test statistic concerning a particular variable of interest exceeded a conventional critical value at a given significance level. In our case a level of $\alpha = 0.05$ was used.

It might be considered to take the individual study as the unit of analysis. However, it must be noted that some studies deal with multiple outcome indicators and/or multiple indicators of concepts relating to teacher and/or school effectiveness. For example, when a (hypothetical) study is using two indicators on the factor time (e.g. time spent on mathematics and time spent on reading) and assesses the impact of each indicator on two outcomes (e.g. mathematic test and reading test), there are four relationships (2 process indicators × 2 outcomes). These four relationships were combined into two relationships by averaging the effects of both indicators for each of the outcomes. The resulting two relationships were examined for their direction and significance and, consequently, these two results (or replications) were included in our final data set. A similar approach was used when one indicator is used in studies carried out in several countries, i.e. the relationship for each country is entered as a separate replication.

The conventional vote-counting procedure has been criticized on several grounds. First, it does not incorporate sample size into the vote. As sample sizes increase, the probability of obtaining statistically significant results increases. Second, the procedure does not allow the researcher to determine which treatment is the best in an absolute sense. Although information is found about the best treatment, it is unknown what the margin of superiority is; it does not provide an effect size estimate. Third, the procedure has a very low power for the range of sample sizes and effect sizes most common in the social sciences. When effect sizes are medium to small, the conventional vote-counting procedure frequently fails to detect the effects. Finally, when a vote-counting procedure is followed, associating counts with moderator variables (study or content characteristics that might influence whether a replication is significant or not) can only be done in a relatively crude way, as compared to procedures in which effect sizes for each replication are calculated.

Given this critique, an approach to calculate quantitative average effect sizes across studies, was used in addition to the vote-counting procedure in the case of the school factors. Moreover, the quantitative approach, to be explained in more detail below, makes use of the fact that effect sizes vary among studies. Analysing this variation makes it possible to establish whether specific study characteristics may account for it. So, for the school factors we used both a vote counting and a quantitative meta-analysis. For the meta-analysis of teaching effectiveness studies, we only used the quantitative technique.

More specifically, a multi-level approach to meta-analysis (Raudenbusch and Bryk 1986; Hox 2002) was applied. In this approach the selected studies are

considered to be a sample from the population of studies; in our case this regards the relationship between specific school/teaching effectiveness indicators and student outcomes. Nested under each study are the secondary units: the schools. Each study can then be viewed as an independent replication. This concept could be used but would not solve the problem of biased estimates due to unidentified dependencies when applying multiple results from one study, e.g. when effects are reported for mathematics and language achievement in one study, while using the same sample of schools and students. To deal with this problem, instead of the two-level model for meta-analysis a three-level model was used, in which the highest level of the studies is referred to as the across-replication level, and the multiple results within a study as the within-replication level. The principal advantages of the statistical meta-analysis employed are threefold: first, the information from each study is weighted by the reliability of the information, in this case the sample size and second, dependencies between within study replications are controlled for. Third, the method applied enables us to examine which study characteristics (or moderators) are responsible for the variation in effect sizes. Further details of the multi-level approach to meta-analysis that was used are given in the technical annex.

Results 1: The Effects of the School-Level Factors

Independent and Dependent Variables Used in the Meta-Analysis

From the thirteen effectiveness-enhancing conditions listed in Table 8.3, ten were included in our current meta-analysis. The selection was motivated by our intention to concentrate on variables that have a meaning at school level, despite the fact that most of them also have an interpretation at class/teacher level. The variables from the list in Table 8.4 that we did not include are those that are intrinsically characteristics of instructional processes, namely *structured instruction*, *independent learning* and *feedback and reinforcement.*

Dependent variables were student outcomes in the cognitive domain, namely student achievement results in mathematics, language and other subjects, including science.

Moderator Variables Used in the Meta-Analysis

As it was stated above, our method allows us to model effect sizes as a function of study characteristics. A first relevant characteristic deals with the question of whether studies have used a language, a mathematics test score or another score to

assess student achievement. This moderator provides insight into the question as to which learning outcomes are most "malleable" by school characteristics. Previous studies (Scheerens and Bosker 1997) suggest that schools have more impact in the area of mathematics than in the area of language. In our study 45.3 % of our data relate to the use of a math test, 33.8 % of all results to a language test.

Apart from examining the impact of the type of test employed, we also investigated the effects of the country in which the study was conducted (the United States of America, the Netherlands or other countries) and the education level or sector in which the study took place (primary or secondary education). Results regarding these study characteristics provide insight into the question of which context is most "susceptible" for school effectiveness indicators. Studies from the past show that, by and large, effect sizes are higher in US schools and in primary schools (Scheerens and Bosker 1997, Chap. 6). In our study 33.5 % of all effect sizes relate to studies conducted in the US, 24.5 % to studies carried out in the Netherlands and 42 % to studies conducted in other countries. With regard to school type 63.7 % of all results relate to studies carried out in primary schools (36.3 % in secondary schools).

The other moderator variables relate to the quality of the studies involved. One of them relates to the issue whether or not studies control for student intake characteristics. Effect sizes are by definition less accurate in case outcomes are not corrected for student intake characteristics. Almost all studies in our database include characteristics such as socio-economic status, age, gender, ethnicity and, in a minority of cases, prior achievement, implying that only in rare cases the dependent variable represents learning gain. In this study 73.2 % of our results were based on outcomes adjusted for covariables. Another feature of school effectiveness research is its reliance on correlational design and consequently on statistical techniques suited to analyse data produced by such a design. With regards to statistical techniques in the nineties, the school effectiveness community witnessed the rise of multi-level analysis, a technique which takes into account the "nested" character of data and therefore yields more precise and accurate results than studies relying on, for example aggregate data. Nowadays, this technique is prevalent in school effectiveness research, although other more traditional techniques are still employed as well. This raises the question of whether studies applying "high quality" techniques such as multi-level analysis yield different results from studies in which more traditional techniques have been used. We therefore included "multi-level/not multi-level" as an additional moderator variable in our analyses. In our study 57.6 % of all results are based upon multi-level techniques, the other 42.4 % on other techniques.

Finally, most of our independent variables have a meaning at school as well as at classroom level and for these variables an additional moderator was included in the analysis. The moderator in question represents the level at which the school effectiveness indicator of interest was measured; was the indicator measured at the class or school level? An example concerns the analysis of data relating to the concept of *monitoring*. This indicator is sometimes measured at the school level

(for example, by investigating whether a monitoring system is used by the school) and sometimes at the class level (for example, by checking the amount of time spent by a teacher in monitoring pupil's progress). The second example is school climate versus classroom climate. For the indicators *consensus and cohesion among staff*, *parental involvement*, *educational leadership* and *differentiation* only school-level information was used, so that the "level" moderator was not applied to these.

To indicate the effect of school effectiveness variables, Fisher's Z transformation of the correlation coefficient was used. Not all studies presented their results in terms of correlations, and therefore all other effect size measures were transformed into correlations, using formulae presented by Rosenthal (1994). For small values of the correlation coefficient, Z_r and r do not differ much, but it should be remembered that all figures presented in the following and indicating effect sizes refer to Z_r. More information about the statistical procedure followed is provided in the appendix.

Vote Counting

The results representing vote counts are shown in Table 8.6.

The results indicate that in most cases relationships between school effectiveness indicators and student outcome variables are not significant. This is the case for about 70 % of all relationships examined in this study. Only in a minority of cases the relationships between school effectiveness indicators and outcome variables are negative. In total about 4 % of all relationships are negatively significant. Around 25 % of all relationship examined are positively significant. The latter figure could lead to a rather pessimistic conclusion, but it is an average, behind which important differences between indicators are hiding. For example, relationships involving the indicator *effective learning time* turn out to be positively significant in 42 % of all cases and in this respect it seems to be the most robust factor. On the other hand, the indicator *consensus and cohesion among staff* is characterized by the fact that only 6 % of all examined relationships are positively significant. Another weak variable in this respect is the variable *educational leadership*. Only in 8 % of all cases positive, significant relationships are found.

A further examination of the vote-counting results indicates that positive findings are mostly found for studies carried out in the United States and for studies using mathematic tests, studies not employing multi-level techniques and for studies using gross, unadjusted outcomes. These results apply for (almost) every school effectiveness indicator examined in this study. More mixed results are found with respect to the moderator variable "type of school". Depending on the school effectiveness indicator, sometimes more positive findings are found in studies conducted in primary schools, other times in studies in secondary schools. The same conclusion can be reached with regard to the moderator "level of measurement" and the question whether or not multi-level analyses were used.

Table 8.6 Vote counts based on replications (%)

	Primary schools			Secondary schools			Subject-math			Subject-language			Subject other than math and language		
	−	n. s.	+	−	n. s.	+	−	n. s.	+	−	n. s.	+	−	n. s.	+
Consensus and cohesion among staff (cooperation)	9	89	2	0	89	11	12	82	6	0	97	3	0	82	18
Orderly climate	0	75	25	6	62	32	1	66	33	0	82	18	16	53	31
Monitoring	5	72	23	8	61	31	4	67	29	6	72	23	15	61	24
Curriculum quality; OTL	3	72	24	0	64	36	0	53	47	6	82	12	0	78	22
Homework	0	92	8	0	55	45	0	80	20	0	92	8	0	67	33
Effective learning time	0	64	36	0	38	62	0	59	41	0	67	33	0	46	54
Parental involvement	3	63	34	4	73	23	5	64	31	4	72	25	0	53	47
Achievement orientation	2	57	41	6	63	30	2	54	44	8	64	28	3	66	31
Educational Leadership	7	84	9	2	94	4	5	88	7	0	98	2	13	71	17
Differentiation	7	69	24	43	57	0	9	66	25	8	71	21	50	50	0

	Country-USA			Country-The Netherlands			Country other than the USA and NLD			Design gross			Design value added		
	−	n. s.	+	−	n. s.	+	−	n.s.	+	−	n.s.	+	−	n. s.	+
Consensus and cohesion among staff (cooperation)	0	83	17	12	84	4	3	97	0	5	95	0	5	88	8
Orderly climate	0	63	37	0	89	11	7	60	32	11	45	43	0	78	22
Monitoring	0	82	18	6	78	16	9	56	35	10	45	45	6	74	20
Curriculum quality; OTL	8	38	54	0	84	16	0	82	18	0	100	0	2	68	29
Homework	0	58	42	0	85	15	0	100	0	0	75	25	0	79	21
Effective learning time	0	60	40	0	63	38	0	52	48	0	50	50	0	60	40
Parental involvement	2	48	51	0	92	8	6	77	17	0	40	60	5	76	19
Achievement orientation	0	36	64	3	87	10	9	69	22	6	51	43	3	66	30
Educational Leadership	5	84	11	10	88	3	2	92	6	8	89	3	4	86	11
Differentiation	0	35	65	0	72	28	16	77	6	0	44	56	10	70	19

	Technique multi-level			Technique not multi-level			School level			Teacher level			Total		
	−	n. s.	+	−	n. s.	+	−	n. s.	+	−	n. s.	+	−	n. s.	+
Consensus and cohesion among staff (cooperation)	4	90	6	6	88	6	5	89	6				5	89	6
Orderly climate	0	89	11	6	54	40	0	74	26	7	62	31	4	68	29
Monitoring	8	74	18	5	56	38	2	68	29	8	67	25	7	67	26
Curriculum quality; OTL	0	72	28	9	64	27	0	82	18	3	66	31	2	70	28
Homework	0	84	16	0	43	57	0	0	100	0	81	19	0	79	21
Effective learning time	0	71	29	0	44	56	0	48	52	0	60	40	0	58	42
Parental involvement	4	82	14	4	39	58	4	65	32				4	65	32

(continued)

Table 8.6 (continued)

	Technique multi-level			Technique not multi-level			School level			Teacher level			Total		
	−	n. s.	+	−	n. s.	+	−	n. s.	+	−	n. s.	+	−	n. s.	+
Achievement orientation	6	83	11	3	41	56	7	59	33	0	63	37	4	61	35
Educational Leadership	1	94	4	8	82	10	5	87	8				5	87	8
Differentiation	11	72	17	0	47	53	9	68	22				9	68	22

− Negative effect, significant at 0.05; *n.s.* Non-significant effect; + Positive effect, significant at 0.05

Multi-level Approach

The results of the multi-level approach to meta-analysis are presented in Tables 8.7 and 8.8. Table 8.7 shows the average effect sizes for all independent variables, and results are generalized over all moderator variables, including assessment domains (mathematics, reading and other).

Table 8.8 shows the impact of the various moderator variables. Results of the quantitative meta-analyses will be presented by considering the effectiveness-enhancing school factors, our "independent" variables, in the order of which they are listed in Tables 8.7 and 8.8.

Table 8.7 Multi-level, empty model

	Number of cases		Variance		
	Across replications	Within replications	Mean effect size	Across replications	Within replications
Consensus and cohesion among staff (cooperation)	28	83	0.019	0.001	0.000
Orderly climate	46	170	0.129***	0.026*	0.008**
Monitoring	43	194	0.061***	0.003	0.021**
Curriculum quality; OTL	25	43	0.145***	0.028	0.007
Homework	21	56	0.073**	0.019	0.000
Effective learning time	30	111	0.147***	0.014**	0.017
Parental involvement	42	142	0.093***	0.018***	0.000
Achievement orientation	50	135	0.141***	0.036***	0.010
Educational Leadership	53	170	0.046*	0.025*	0.000
Differentiation	30	107	0.017	0.021***	0.008

Effect-sizes marked as * are significant at the 0.10 level, those marked as ** are significant at the 0.05 level, and those marked as *** are significant at the 0.001 level

Table 8.8 Multi-level, with moderators as predictors

										Variance	
	Intercept	Secondary	Arithmetic/Math	Language	USA	The Netherlands	Value added	Not multi-level	Class/teacher level	Across replications	Within replications
Consensus and cohesion among staff (cooperation)	−0.058***	0.065***	−0.006	0.004	0.053***	0.031*	0.032**	0.004	–	0.000	0.000
Orderly climate	0.135	0.025	−0.004	0.001	0.082	0.005	−0.076	0.017	−0.009	0.023**	0.008**
Monitoring	0.121**	−0.070**	0.106*	0.117**	−0.116**	−0.098***	−0.084	0.088***	−0.052**	0.001	0.019**
Curriculum quality; OTL	−0.047	−0.066	0.151	0.008	−0.047	−0.007	0.110	0.199	0.013	0.012**	0.006**
Homework	0.233	−0.055	0.036	0.020	0.335***	0.376***	−0.410***	0.087	−0.166**	0.000	0.000
Effective learning time	0.191***	−0.185**	−0.039	0.058	−0.092	−0.145*	0.039	0.210***	−0.090	0.002	0.019*
Parental involvement	0.213***	−0.005***	−0.010	−0.013	0.114*	−0.028	−0.136**	−0.105	–	0.019***	0.000
Achievement orientation	0.202***	−0.063	−0.022	−0.027	0.070	−0.154**	−0.180	0.000	0.047	0.028***	0.010
Educational leadership	0.052	−0.002	0.009	0.011	0.050	−0.095	0.012	−0.018	–	0.022**	0.000
Differentiation	−0.085	−0.067	−0.035	−0.007	0.245**	0.052	0.046	0.232***	–	0.006**	0.007

Effect sizes marked as * are significant at the 0.10 level, those marked as ** are significant at the 0.05 level, and those marked as *** are significant at the 0.001 level

Cooperation

The results of the meta-analysis for the factor *consensus and cohesion among staff* (*cooperation*) and its impact on pupil achievement show that in total 28 studies were included, some of which contained multiple results, leading up to a total number of 83 within replications.

The estimated mean effect size of cooperation across all studies equals a Fischer's Z value of 0.019. The estimated variance across all studies (both within and across replications) is 0.001. This indicates that the 95 % prediction interval around the means ranges between $Z_r = -0.043$ and $Z_r = 0.081$.

The prediction interval, in contrast to the confidence interval, describes the distribution based on the estimates. The confidence interval gives only information on the degree of precision with which the mean of that distribution is estimated.

The results of the analysis trying to predict differences between effect sizes with moderators such as subject matter, sector, study design and others indicate that some of the moderators have a significant relationship with the effect size. Studies carried out in the USA and the Netherlands, studies carried out in secondary schools and studies employing a value-added design show significantly higher effect sizes. These results do not change the overall conclusion that cooperation among teachers appears to be an insignificant variable in explaining variation in pupil achievement. For example, controlling for other study characteristics, US studies have an average effect size around zero (−0.058 + 0.053). For the Netherlands this figure is −0.027 (−0.058 + 0.031).

Orderly Climate

The estimated mean effect size of orderly climate is 0.129, which is significant at the 1 % level. The mean effect size is based on total 46 studies, most of them with multiple results. The total of within replications is 170. The estimated variation across all studies (both across and within replications) equals 0.034. This indicates that the 95 % prediction interval around the mean effect size is between $Z_r = -0.231$ and $Z_r = 0.489$.

The results of the analyses trying to predict differences between effect sizes show that none of the moderators has a significant relationship with the mean effect size. This means, for example that there is no difference between studies measuring this concept at the school level or studies measuring it at the class level; the effect size in both types of studies is equal.

Monitoring

The estimated effect size of monitoring across all 25 studies involved in our analyses is $Z_r = 0.06$ which is significant at the 1 % level. The estimated variance across all studies (both within and across replications) is 0.024, indicating that the

95 % prediction interval around the mean effect size runs from $Z_r = -0.243$ to 0.3625. The results of the analyses aimed at establishing the relationship between effect sizes with study characteristics show that many moderators have a significant relationship with the effect size. Some of them have comparably a rather strong positive relationship with the effect size, others a rather strong negative relationship. For example, when a language test is the outcome variable, the effect size is about 0.12 higher, and when an arithmetic or mathematics test is used the effect size is about 0.11 higher, than in cases where other outcome variables were used. This implies that the mean effect size for monitoring is strongly diminished when outcomes in other subject are included, whereas effect sizes for language and mathematics are relatively high, 0.18 and 0.17, respectively. A smaller difference (0.09) in effect size is noted with respect to studies that do not use multi-level modelling, as compared to studies that do use multi-level modelling. On the other hand, effect sizes turn out lower for studies carried out in the USA compared to all other countries. The same results apply to studies carried out in the Netherlands. On average, effect sizes for studies carried out in the Netherlands and the USA are around zero. When monitoring is measured at school level it has a higher effect size than in cases where it is measured at classroom level (0.05 higher). Finally, studies conducted in secondary schools show lower effect sizes than studies carried out in primary education (difference 0.07).

Curriculum Quality

The curriculum quality concept includes three variables; opportunity to learn, effective learning time and homework.

The analysis concerning *opportunity to learn* involves 25 studies with 43 results in total. The mean effect size is 0.145 which is significant at the 1 % level. The 95 % prediction interval around the means ranges between $Z_r = -0.222$ and 0.512. The analysis reveals further that there is hardly any variance among studies with regard to their effect sizes. Not surprisingly, there are no significant relationships between the moderators and the effect size.

With regard to *homework* 21 studies were analysed involving a total of 56 results. The estimated mean effect size is 0.073. This indicates that the 95 % prediction interval ranges between $Z_r = -0.197$ and 0.343. Although the variance in effect appears relatively small, there are important differences between studies and the effect sizes they yield. This regards, first of all differences between countries. Studies conducted in both the US and the Netherlands yield effect sizes which are much higher than the effect sizes yielded by studies carried out in other countries. US studies have an effect size which differs 0.345 from all other studies, while Dutch studies differ 0.376 from all other studies. Moreover, studies employing multi-level techniques produce much lower effect sizes. Finally, studies measuring the concept at the class level (in fact 98 % of all studies) yield significantly lower effect size than studies measuring the concept of homework at the school level.

The estimated effect size of *effective learning time* equals 0.147 (significant at the 1 % level). This indicates that the 95 % prediction interval ranges between $Z_r = -0.0197$ and 0.491. The analysis relating moderators to the effect size indicate that studies carried out in primary schools show significantly lower effect sizes (0.19), while studies employing other than multi-level techniques yield significantly higher effect sizes (a difference of 0.21). Finally, there is also a difference between countries. Studies carried out in the Netherlands come up with significantly lower effect sizes (−0.145). On average, the effect size of Dutch studies is about 0.05 (0.191 − 0.145).

Parental Involvement

The analyses concerning parental involvement involve 42 studies, again with most of them having multiple results. In total there are 142 replications within the studies. The estimated effect size of parental involvement in all studies is $Z_r = 0.093$, which is significant at the 1 % level. The 95 % prediction interval around the means ranges between $Z_r = -0.169$ and 0.355.

The data also show significant variation in effect sizes. The most important moderators in this respect are, respectively, whether or not the study involved controls for student characteristics affecting learning achievement and the country in which the study has been carried out. Not surprisingly, with regard to the former, studies taking into account student characteristics show significantly lower effect sizes (difference in coefficient of −0.14). With regard to the latter, effect size of studies carried out in the US is significantly higher than effect size of studies carried out in all other countries (difference in coefficient of 0.11); controlling for the impact of other moderators, the effect size in US studies is on average 0.327 (0.213 + 0.114).

Achievement Orientation

The estimated mean effect size for achievement orientation is 0.147, which is significant at the 1 % level. The 95 % prediction interval around the means ranges between $Z_r = -0.279$ and 0.561. The figures presented are based on 30 studies containing 81 results.

Once again the data indicate that there is significant variation in effect sizes across studies. However, only one moderator is of significance in this respect. Studies conducted in the Netherlands have significantly lower effect sizes than studies carried out in other countries (−0.15). Moreover, an interesting fact is that it does not seem to matter at which level this concept is measured. There is no significant difference between studies measuring this concept at the school level and studies measuring this concept at the class level.

Educational Leadership

Another frequently studied school effectiveness indicator is educational leadership (53 studies with 170 results). The mean effect size in this case is 0.046. This figure is significant at the 10 % level. The 95 % prediction interval around the means ranges between $Z_r = -0.263$ and 0.355. The analyses cannot detect any significant relationship between the moderators distinguished in this study and the effect size, although studies' effect sizes vary significantly around the mean.

Differentiation

The last concept investigated in this study is differentiation. The analysis concerning this concept involves 30 studies with in total 107 different results. The mean effect size found is 0.017, a figure which does not deviate significantly from zero. The 95 % prediction interval lies between $Z_r = -0.317$ and 0.351. The results of the analysis examining the variation in the effect size show that two moderators are important. Effect sizes are significantly higher in US studies than in studies conducted in all other countries (a difference of 0.245) and in studies using other techniques than multi-level techniques (a difference of 0.232).

Conclusion of the Results on the School-Level Variables

Overviewing our quantitative results we can observe that the effect sizes range between 0.017 and 0.147. In terms of Cohen's *d* (which is approximately twice the size of the correlation coefficient) this means that the results vary from negligible to small.

In this respect they resemble the results of a previous meta-analysis presented by Scheerens and Bosker (1997). The results are also similar in the sense that the effect sizes found in this study for the different effectiveness indicators are comparable to the ones found previously. The biggest differences are found with respect to *parental involvement* (now 0.09, then 0.13), *effective learning time* (now about 0.15, then 0.19), *monitoring* (now 0.06, then 0.14) and *curriculum quality (opportunity to learn)* (now 0.13, then 0.08). The conclusion with respect to monitoring should be modified, however, since effect sizes in important subject matter areas such as language and mathematics are in the order of 0.18 and 0.17 in our current analysis.

Another conclusion is that the quantitative results, i.e. the calculated effect sizes and their relationships with the moderator variables, are not as clear-cut as the vote-counting results. The results based on the vote counting showed a clearer pattern concerning particular moderators, for example studies conducted in primary schools turned out to have more positive, significant relationships than studies carried out in secondary schools. The conclusions with regard to results based on the multi-level approach to meta-analysis are somewhat less consistent across the independent variables, which makes it difficult to draw unequivocal conclusion

about the impact of specific moderators. Yet, our results indicate that, for practically all variables, effect sizes are smaller when outcomes are adjusted for student background characteristics, and for all but two variables effect sizes are smaller when multi-level analyses are applied. There is also a relatively consistent slightly higher effect size for studies carried out in primary, as compared to studies conducted in secondary schools. The picture is less clear-cut for the moderator variables subject matter area and country.

Discussion of the School-Level Results

According to Cohen's standards for interpreting effect sizes,[1] our results on school effectiveness indicators should be interpreted as negligible to small. It should be noted however, that several authors argue that Cohen's standards are to be considered as too conservative, and do not match the practical significance of malleable school variables. Richard et al. (2003; cited by Baumert et al. 2006) found a mean correlation of $r = 0.21$ in their meta-analysis of meta-analyses in social psychology, and proposed a modification of Cohen's classification, considering a correlation of 0.30 to indicate a large effect (p. 339). Baumert et al. (2006) propose the learning gain during one school year as a realistic standard to express effects of schooling. They cite several studies that indicate that this learning gain has the magnitude of about $d = 0.30$. These authors also discuss a method to compute effect sizes developed by Tymms et al. (1997), which, when applied to a practical example, suggests that effect sizes of about $r = 0.15$–0.20 (small to medium, according to Cohen's standards) would equal the learning gain in one school year, which they consider as an effect of great practical relevance. Seen in this light the effect sizes that we found for a number of school effectiveness indicators (in particular school climate, curriculum quality, learning time and achievement orientation) should be upgraded in their rating for practical significance.

Among the set of school effectiveness indicators that were studied the curriculum-related and climate-related factors showed the largest effects. *Opportunity to learn* and *leaning time* had effect sizes of 0.15; whereas *orderly climate* and *achievement orientation* had effect sizes of 0.13 and 0.14, respectively. The relative importance of the curriculum variables underlines the importance of the content dimension in schooling. The time factor is interpreted in the sense of the temporary engagement with content, and, in this way, as a dimension of the implemented curriculum. The results on *homework* can be given a similar interpretation, where the effect size for homework was 0.07. The realization that content and exposure to content matters could be interpreted as supporting the view that proactive structuring of content, as in externally developed curricula and lesson

[1]According to Cohen (1988), small effects are in the order of $r = 0.10$, medium effects $r = 0.30$ and large effects $r = 0.50$ or higher.

plans, has a rightful place among school improvement strategies. This result speaks to the debate concerning school-based, "bottom up" school improvement strategies versus the implementation of external curricula. The former approach has been the preferred approach among scholars in the field of educational change (cf. Miles 1998) but has been criticized, among others by Slavin (2000), and Muijs and Reynolds (2001) who describe the bottom up approach as "the ownership paradigm", in which the "re-invention of the wheel" by individual schools is put down as an inefficient approach. A similar line of argumentation, favouring externally developed curriculum material, is used with respect to the approach followed in Comprehensive School Reform Programs in the USA (Borman et al. 2004).

The relatively high effect sizes concerning an *orderly school climate* are in line with results from large-scale international assessment studies, like OECD's PISA. More in depth analyses of these results (Luyten et al. 2005), however, indicated that the climate effects were heavily confounded with school composition, in the sense of school average socio-economic status (SES) of the students. More specifically these results showed that schools with a better climate were more likely to have higher level SES composition.

The second climate factor, *achievement orientation* is based on variables like: clear focus on mastering basic subject, high expectations of students' achievement and record keeping of students' achievement. High expectations reflect an active, optimistic attitude that seeks to get the best out of all students, and is related to the personality characteristic of internal locus of control. At the same time measures of high expectations might express a more reactive attitude, in which relatively high achievement is more like a cause, rather than an effect of high expectations.

Two variables that should be considered of high policy relevance in effective schooling, *monitoring* and *educational leadership* came out as having very small average effect sizes (0.06 and 0.05, respectively). The evaluation and feedback mechanism is considered as a promising lever for organizational learning and school improvement, an expectation that is at least reasonably met for language and mathematics outcomes (effect sizes of 0.18 and 0.17, respectively). Scheerens and Bosker (1997) report an average effect size for monitoring of 0.15. In evaluation studies concerning types of school evaluation and monitoring results show a mixed pattern as well. Schildkamp reports relatively disappointing results of evaluations of school self-evaluation programmes (Schildkamp 2007). Research results on the impact of system-level accountability policies (Carnoy et al. 2003), however, indicate that the combination of a high internal evaluative potential of schools and a context of high stakes external accountability policy is effective in enhancing student performance.

The effect size for *educational leadership* (0.05) confirms a similar effect size as reported in Scheerens and Bosker (1997). More in depth analysis by Witziers et al. (2003) focused on indirect effects of educational leadership, where the interesting question is the one about the identification of variables that mediate the effect of leadership. Their results, and those of later studies, provide little consistency between studies, concerning the intermediary variables that were identified. Table 8.9 provides an overview.

Table 8.9 Intermediary variables in studying indirect effects of school leadership

Reference of study	Significant intermediary variables
Hallinger and Heck (1998)	Learning climate
	Principal's instructional efforts
Hallinger et al. (1996)	A clear school mission
	Students' opportunity to learn
	Teachers' expectations
Hill et al. (1995)	Teacher–student interactions
	Professional climate
Bosker et al. (2000)	Teachers' job satisfaction
	Teachers' achievement orientation
	Evaluation and feedback practices
Kythreotis and Pashiardis (2006)	Teachers' commitment to the school
	Teachers' academic emphasis
	Personal achievement goal orientations
	Classroom performance-goal structure

The school effectiveness indicator *cooperation and consensus* is a factor that makes perfect practical sense and has an important place in conceptual models of school effectiveness (e.g. the model developed by Creemers; cf. Creemers and Kyriakides 2006). It comes out weak in our current and previous meta-analyses 0.02 and 0.03, respectively. This low effect may be due to the rather superfluous way in which this variable is often measured, for example in terms of the frequency of staff meetings.

Parental involvement had a small effect (0.09) as compared to Cohen's categorization, but, given the consideration mentioned at the beginning of this discussion, this might still be of practical significance (Cohen 1977). *Differentiation* had a negligible effect (0.02), but it should be noted that this factor, measured at school level refers to school-level policies, and would potentially have higher impact when studied at classroom level.

A final issue, of a more methodological nature, has to do with strengths and weaknesses of a meta-analysis in which the effect of each relevant factor is estimated separately. The main advantage is that this approach attempts to show what each and every factor is "worth" in its association with student achievement. A major disadvantage is the fact that the approach does not take the intercorrelations between the factors, nor relevant contextual variables, nor intermediary variables (particularly classroom-level instructional processes) into consideration. In this sense, in order to give a more complete overview of the knowledge base from school effectiveness research, additional review is needed of studies that have examined more complex configurations of the factors that were dealt with as discrete, independent factors in our analysis. Four kinds of studies should be mentioned, that would be complimentary to our approach:

(a) Studies that investigate indirect effects; the examples that were presented in the above, concerning educational leadership, illustrate this approach.
(b) Studies that have attempted to model alternative specifications of across level (e.g. school, classroom, individual student) relationships (e.g. Bosker and Scheerens 1995; Hofman 1995; Reezigt et al. 1999). Such studies are sparse and their potential usually limited by data constraints.
(c) Studies that are driven by elaborate conceptual models of school effectiveness, e.g. Creemers and Kyriakides (2006), Kyriakides (2005). The same qualification applies as for the previous category.
(d) Evaluations and meta-evaluations of Comprehensive School Reform Programs, which include all factors that were studied in our meta-analyses (and more; comprising also instructional variables and above school policies). The positive outcomes of these studies, (e.g. Borman et al. 2004) present probably the most robust empirical support for the "effective school model", so far. The more so, because the quasi-experimental design of the evaluations is one step further in allowing for causal interpretation of the research findings, than is the case for the mostly survey based, non-experimental nature of the typical research study on which our analyses are based.

Results 2: The Effects of the Teaching Variables

The first set of results on the teaching variables shows the mean effect sizes, computed as the association of the teaching variables with cognitive learning outcomes. As explained in the method section, the effect sizes are expressed as Fischer-*Z* coefficients, which can be interpreted as correlations. As indicated in the paragraph on the definition of key variables, two categorizations were used that depended on different aggregations of 46 specific variables. The original categorization was based on the chapter by Seidel and Steen, in Scheerens et al. (2005), and has 15 teaching dimensions. The alternative categorization uses a total of six broader categories. Effect sizes are shown for the two aggregate categorizations and for each of the 46 basic variables. Effect sizes for the 15 dimensions are shown in Tables 8.10 and 8.11.

The second set of results shows the degree to which effect sizes depend on study characteristics such as: study design (experimental or non-experimental), school level (primary or secondary schools), subject matter area (arithmetic/mathematics), country, whether or not multi-level analyses were applied, and whether or not outcomes were adjusted for relevant student background characteristics.

Table 8.10 Mean effect sizes for the 15 main areas as defined by Seidel and Steen (2005)

Area		Mean eff.	St. error	*p*	Count
1	Learning time	0.095	0.013	0.000	173
2	Classroom organization	0.075	0.019	0.000	62
3	Learning environment	0.129	0.012	0.000	138
4	Clear and structured	0.126	0.016	0.000	134
5	Activating	0.123	0.012	0.000	179
6	Learning strategies	0.213	0.018	0.000	103
7	Challenge	0.130	0.013	0.000	180
8	Support	0.108	0.017	0.000	73
9	Feedback	0.056	0.019	0.004	106
10	Evaluation	0.086	0.031	0.006	46
11	Teacher characteristics	0.146	0.039	0.000	26
12	Adaptive teaching	0.066	0.027	0.013	41
13	Practice	-0.080	0.030	0.007	27
14	Material	0.015	0.016	0.033	28
15	Integrative approaches	0.089	0.022	0.000	90

Table 8.11 Mean effect sizes, with 95 % confidence intervals

Area		Mean eff.	St. error	Lower	Upper	Count
1	Learning time	0.095	0.013	0.069	0.121	173
2	Classroom organization	0.075	0.019	0.038	0.111	62
3	Learning environment	0.129	0.012	0.105	0.153	138
4	Clear and structured	0.126	0.016	0.095	0.157	134
5	Activating	0.123	0.012	0.098	0.147	179
6	Learning strategies	0.213	0.018	0.177	0.248	103
7	Challenge	0.130	0.013	0.105	0.155	180
8	Support	0.108	0.017	0.075	0.140	73
9	Feedback	0.056	0.019	0.018	0.094	106
10	Evaluation	0.086	0.031	0.025	0.148	46
11	Teacher characteristics	0.146	0.039	0.070	0.222	26
12	Adaptive teaching	0.066	0.027	0.014	0.118	41
13	Practice	−0.080	0.030	−0.138	−0.021	27
14	Material	0.015	0.016	−0.015	0.046	28
15	Integrative approaches	0.089	0.022	0.046	0.131	90

Mean Effect Sizes

The highest effect sizes are teaching learning strategies (0.21), teacher characteristics, such as high expectations and constructivist beliefs about teaching (0.15), a cognitively challenging teaching approach (0.13), a learning environment that is

orderly and characterized by low achievement pressure (0.13), a clear and structured teaching approach (0.13) and "active" teaching, characterized by a variety of didactic approaches (0.12). The coefficient for support was 0.11. The only negative coefficient was found for specific teaching practices, particularly drill and repetitions, while the coefficient for the quality of teaching material, was quite low (0.02). The other areas show effect sizes in the range between 0.06 and 0.09.

The information in Table 8.12 gives more detail, as it shows the coefficients for each of the 46 basic variables, within the 15 areas.

Table 8.12 Mean effect sizes for the 46 variables, categorized according to the 15 areas defined in Seidel and Steen (2005)

Sub-area	Mean eff.	Count	Cat.
1.1 Time on task	0.125	86	II
1.2 Opportunity to learn	0.118	32	I
1.3 Homework	0.041	51	II
1.4 Mastery learning	0.047	4	III
2.1 Classroom management	0.088	36	II
2.2 Discipline	0.070	20	II
2.3 Control	0.018	17	II
3.1 Classroom climate	0.125	107	V
3.2 No achievement pressure	0.151	29	II
3.3 Mastery orientation	−0.005	2	III
3.4 No performance orientation	0.120	2	II
4.1 Structured/direct teaching	0.107	76	III
4.2 Goal-directed/clear	0.222	36	III
4.3 Teacher demonstration	0.014	17	III
4.4 Teaching basic skills	0.073	17	III
5.1 Cooperative	0.204	49	IV
5.2 Situated/discovery	0.155	3	IV
5.3 Peer tutoring	0.218	53	IV
5.4 Student work	0.059	36	IV
5.5 Individual work	−0.009	39	IV
5.6 Student discussions	0.043	8	IV
6.2 Metacognitive	0.244	35	IV
6.3 Scientific inquiry	0.197	32	IV
6.5 Organizing methods	0.000	2	IV
6.7 Reading/writing	0.210	34	IV
7.1 Cognitive activation/understanding orientation	0.182	67	IV
7.2 Active student engagement	0.042	63	IV
7.3 Authentic contexts/relevance	0.160	47	IV
7.4 Language level	0.029	7	II
7.5 Representation formats	0.385	4	II

(continued)

Table 8.12 (continued)

Sub-area	Mean eff.	Count	Cat.
8.1 Quality of interactions/teacher support	0.108	73	V
9.1 Feedback/frame of reference/monitoring	0.056	106	VI
10.1 Assessments/tests	0.086	46	VI
11.1 High expectations	0.124	22	II
11.2 Constructivist beliefs about learning	0.354	4	IV
12.1 Various teaching methods	0.124	2	II
12.2 Adaptive teaching	0.036	27	II
12.3 Open tasks/choice	0.090	4	II
12.4 Student prerequisites	0.178	7	II
13.1 Drill/repetition	−0.078	17	III
13.2 Application	−0.057	19	III
14.1 Textbooks	0.039	6	I
14.2 Media	0.012	27	I
15.1 Constructivist	0.039	52	IV
15.2 Inductive	−0.197	5	IV
15.3 Concept-oriented	0.257	33	IV

The fourth column indicates the aggregation of these variables according to six areas (Annex 4.2 provides 95 % confidence intervals for the mean effect sizes included in Table 8.12)

Some of the variables that show relatively high coefficients depend on just a few replications (e.g. representation formats) and should therefore be taken with a lot of reservation. Teaching metacognitive strategies has a relatively high coefficient with a fair number of replications (0.24 with 35 replications). Peer tutoring, cooperative learning and instruction that is clearly goal directed have relatively high coefficients as well (0.22, 0.20 and 0.21, respectively). Relatively high coefficients are also found for subject specific learning strategies, like scientific inquiry and reading and writing, 0.20 and 0.21, respectively.

It is worth noting that the two variables that represent achievement press in teaching have negative associations with cognitive achievement (no achievement pressure and no performance orientation have positive associations of 0.15 and 0.12, respectively).

Next, results are presented according to the more synthetic categorization, explained in a previous section. This second categorization uses six more general categories. Table 8.13 shows the coefficients for the main areas, whereas Table 8.14 shows the coefficients for the 46 variables, now ordered according to the 6 areas; of course this table contains the same basic information as Table 8.12, above, based on 15 areas.

The results in Table 8.13 show that the coefficients have, to some degree flattened out, as compared to the coefficients in the 15 area categorization. Interestingly, the coefficient for constructivist-oriented teaching strategies is higher than the coefficient for a more structured teaching approach. Yet the difference is

Table 8.13 Categorization in terms of six more general areas

Category		Mean eff.	St. error	*p*	Count
I	Curricular	0.077	0.023	0.001	61
II	Teacher-orchestrated classroom management	0.095	0.010	0.000	304
III	Teaching strategy (structured, direct, mastery …)	0.087	0.015	0.000	165
IV	Teaching strategy (constructivist-oriented …)	0.135	0.008	0.000	542
V	Climate, support, positive interactions	0.117	0.011	0.000	180
VI	Feedback/monitoring/assessment/tests	0.065	0.017	0.000	152

not enormous and the more "teacher centred" areas, like teacher-orchestrated classroom management and a supportive classroom climate, show, next to structured teaching, coefficients that, in this context, should be considered as relevant (see the discussion on the interpretation of effect sizes in the first part of this chapter about school-level factors).

In Table 8.14 the effect sizes of the underlying variables of the six categories are shown.

With respect to the effect sizes Table 8.14 does not provide new information. When comparing the consistency of effect sizes within areas, the classification in Tables 8.13 and 8.14 is not superior to the one based on the 15 areas (Tables 8.10, 8.11 and 8.12); in both classification there is a sizeable amount of variation in effect sizes among the variables that are brought together under one heading. In Table 8.14 the consistency in category IV (constructivist-oriented teaching strategy) is to some extent fair, but the result for variable 15.1 (a constructivist teaching orientation) deviates considerably, while it stands for a teaching approach that has the same label as the term that is used to indicate main area IV. Perhaps this result underlines the fact that of the different components of a constructivist teaching approach particularly the active teaching of metacognitive strategies, as well as certain "social" teaching forms (cooperative learning and peer tutoring) appear to be effective.

The Impact of Study Characteristics on Effect Sizes

For a subset of studies, namely those that were collected over the period between 1995 and 2005, a number of study characteristics were analysed, in order to trace possible method and context effects.[2] This was done by means of multi-level

[2]The fact that a subset was used explains that the number of replications for the 6 areas in Table 8.15 is generally less than in Table 8.13.

Table 8.14 Main effect sizes for the 46 variables categorized according to six areas

Subarea	Mean eff.	Count	Cat.
1.2 Opportunity to learn	0.118	32	I
14.1 Textbooks	0.039	6	I
14.2 Media	0.012	27	I
1.1 Time on task	0.125	86	II
1.3 Homework	0.041	51	II
2.1 Classroom management	0.088	36	II
2.2 Discipline	0.070	20	II
2.3 Control	0.018	17	II
3.2 No achievement pressure	0.151	29	II
3.4 No performance orientation	0.120	2	II
7.4 Language level	0.029	7	II
7.5 Representation formats	0.385	4	II
11.1 High expectations	0.124	22	II
12.1 Various teaching methods	0.124	2	II
12.2 Adaptive teaching	0.036	27	II
12.3 Open tasks/choice	0.090	4	II
12.4 Student prerequisites	0.178	7	II
1.4 Mastery learning	0.047	4	III
3.3 Mastery orientation	−0.005	2	III
4.1 Structured/direct teaching	0.107	76	III
4.2 Goal-directed/clear	0.222	36	III
4.3 Teacher demonstration	0.014	17	III
4.4 Teaching basic skills	0.073	17	III
13.1 Drill/repetition	−0.078	17	III
13.2 Application	−0.057	19	III
5.1 Cooperative	0.204	49	IV
5.2 Situated/discovery	0.155	3	IV
5.3 Peer tutoring	0.218	53	IV
5.4 Student work	0.059	36	IV
5.5 Individual work	−0.009	39	IV
5.6 Student discussions	0.043	8	IV
6.2 Metacognitive	0.244	35	IV
6.3 Scientific inquiry	0.197	32	IV
6.5 Organizing methods	0.000	2	IV
6.7 Reading/writing	0.210	34	IV
7.1 Cognitive activation/understanding orientation	0.182	67	IV
7.2 Active student engagement	0.042	63	IV
7.3 Authentic contexts/relevance	0.160	47	IV
11.2 Constructivist beliefs about learning	0.354	4	IV
15.1 Constructivist	0.039	52	IV

(continued)

Table 8.14 (continued)

Subarea	Mean eff.	Count	Cat.
15.2 Inductive	−0.197	5	IV
15.3 Concept-oriented	0.257	33	IV
3.1 Classroom climate	0.125	107	V
8.1 Quality of interactions/teacher support	0.108	73	V
9.1 Feedback/frame of reference/monitoring	0.056	106	VI
10.1 Assessments/tests	0.086	46	VI

analyses as were described in the methods section. In order to have analyses with sufficient statistical power the analyses were limited to the six-category categorization of the teaching variables. The overall results are shown in Table 8.15, and the results for the different moderator variables in Tables 8.16 and 8.17. Comparison between Tables 8.13 and 8.15 shows that the mean effect sizes based on the more limited number of replications in Table 8.15 are slightly higher for category I (curriculum); and slightly lower for all other categories, except category IV (constructivist teaching strategies); for the latter category the coefficient and the number of replications are practically the same.

The results shown in Tables 8.16 and 8.17 indicate that the six teaching factors have generally higher effect sizes in primary as compared to secondary education; the curriculum factor has a significantly higher effect size for mathematics; as compared to studies in all other domains, while feedback has significantly higher effect sizes for both mathematics and science as compared to studies based on outcomes in all other subject matter areas. When studies use stricter methodological

Table 8.15 Multi-level, empty model

	Number of cases			Variance	
	Across replications	Within replications	Mean effect size	Across replications	Within replications
I Curricular	19	50	0.110**	0.044	0.000
II Teacher-orchestrated classroom management	46	246	0.081***	0.011***	0.011**
III Teaching strategy (structured, direct, mastery …)	39	165	0.078***	0.008***	0.018***
IV Teaching strategy (constructivist-oriented …)	70	541	0.139***	0.013***	0.011***
V Climate, support, positive interactions	28	156	0.091***	0.007***	0.006**
VI Feedback, monitoring, assessment, tests	21	119	0.060**	0.004	0.023**

Effect sizes marked as * are significant at the 0.10 level, those marked as ** are significant at the 0.05 level, and those marked as *** are significant at the 0.001 level

Table 8.16 Multi-level, with moderators as predictors

	Intercept	Secondary	Arithmetic/Math	Language	Value added	Not multi-level
I Curricular	0.266*	−0.239***	0.215**	−0.016	–	0.156*
II Teacher-orchestrated classroom management	0.070	−0.078	0.007	−0.012	0.014	0.089**
III Teaching strategy (structured, direct, mastery …)	0.177***	−0.111***	0.027	0.026	−0.074	0.047
IV Teaching strategy (constructivist-oriented …)	0.111*	−0.081**	0.009	0.011	−0.053	0.126***
V Climate, support, positive interactions	0.121**	−0.047	−0.011	0.005	−0.053	0.046
VI Feedback, monitoring, assessment, tests	0.106**	−0.155***	0.136**	0.143**	−0.144***	0.060***

Effect sizes marked as * are significant at the 0.10 level, those marked as ** are significant at the 0.05 level, and those marked as *** are significant at the 0.001 level

Table 8.17 Multi-level, with moderators as predictors (continuation of Table 8.16)

					Variance	
	USA	The Netherlands	Germany	Experimental	Across replications	Within replications
I Curricular	−0.231***	−0.072	0.002	−0.097	0.010**	0.000
II Teacher-orchestrated classroom management	−0.041	−0.026	0.038	0.004	0.008**	0.011**
III Teaching strategy (structured, direct, mastery …)	−0.052	0.028	−0.017	0.016	0.004	0.017***
IV Teaching strategy (constructivist-oriented …)	−0.001	−0.038	0.014	0.042	0.008***	0.010***
V Climate, support, positive interactions	−0.020	−0.020	0.113***	−0.080**	0.003**	0.006**
VI Feedback, monitoring, assessment, tests	−0.093***	−0.084**	0.123***	−0.055	0.000	0.020**

controls, like a value-added design or application of multi-level methods, the effect sizes tend to be smaller. Country effects are found for the curriculum factor and the feedback factor having significantly lower effect sizes in the USA as compared to all other countries, the feedback factor showing a significantly lower effect size in the Netherlands as well, while the supportive climate and feedback factor show significantly higher effect sizes in Germany.

Contrary to expectations effect sizes for experimental studies are not significantly higher than for correlational studies; for the supportive climate factor effect sizes in experimental or quasi-experimental studies are even significantly lower than for correlational studies.

We had expected significant positive effect of experimental and quasi-experimental studies, particularly in the domain of teaching strategies, factors

Table 8.18 Comparison of effect sizes of teaching factors for (quasi-) experimental versus correlational studies

Category	(Quasi-) experimental		Correlational	
	Mean eff.	Count	Mean eff.	Count
I Curricular	0.08	9	0.06	41
II Teacher-orchestrated classroom management	0.11	36	0.09	210
III Teaching strategy (structured, direct, mastery …)	0.13	66	0.06	99
IV Teaching strategy (constructivist-oriented …)	0.16	337	0.10	204
V Climate, support, positive interactions	0.04	31	0.14	125
VI Feedback/monitoring/assessment/tests	0.05	14	0.08	105

III and IV, because these studies lend themselves more to well-controlled intervention studies than variables that reflect more general conditions such as classroom climate. Although the results show a positive trend, the influence of the moderator variable in question did not reach statistical significance in the multi-level analyses. To explore this issue further, effect sizes for the six teaching factors were pairwise compared between (quasi) experimental and correlational. Results are shown in Table 8.18.

With respect to the teaching strategy variables effect sizes for (quasi) experimental studies tend to be higher than for correlational studies (0.13 vs. 0.06 and 0.16 vs. 0.10). Although these differences are in the expected direction, they are not dramatically high. As one might argue that correlational studies that reflect variance in "real life" situations may have generally higher ecological validity than specific experimental intervention studies, the finding that the effect sizes for the former are not that much lower than for studies higher in internal validity can be interpreted as positive evidence for the robustness of the factors in question.

Conclusions

When comparing the effect sizes of the teaching effectiveness factors to those of the school effectiveness factors, reported in earlier sections of this chapter, it can be noted that the range of effect sizes for the teaching-level factors is between −0.080 and 0.213, whereas the range for the school-level factors was between 0.017 and 0.147. On average the teaching effects are somewhat larger than the effects of the school-level variables; this confirms the often established result that independent variables more proximal to student learning have higher impact on student achievement than more distal variables, i.e. those defined at school or educational system level (cf. Wang et al. 1993).

As far as the impact of moderator variables on the effect sizes are concerned the only relatively consistent trend across teaching factors was found for school sector: as in the case of the school-level variables the effect sizes were generally higher in primary than in secondary schools. Since the study of teaching factors lends itself more easily to the application of experimental and quasi-experimental results, and one would expect higher effects in better controlled intervention studies it was expected to find a "design effect" of higher effect sizes in experimental studies as compared to correlational studies. Also for teaching strategy variables (i.e. those that either represent clear and structured "direct teaching" or more constructivist-oriented teaching strategies, like teaching metacognitive strategies and cognitive activation) the trend was in this direction, the effect of this moderator variable, however, did not reach statistical significance. In a substantive way this outcome should be seen as confirming the robustness and ecological validity of the teaching strategy variables in question, as they "show up" regardless of experimental or correlational study design.

Discussion of the Results of the Meta-Analysis on the Teaching Variables

Effect Sizes

A similar discussion as was conducted with respect to the effect sizes of the school variables can be applied to the teaching variables. The average effect size of the school variables was about $r = 0.08$, whereas the average effect size for the teaching variables was about $r = 0.10$. Although the latter is slightly higher, both are in the realm of small effects according to Cohen, but "medium" according to the reasoning of some educational researchers (e.g. Tymms et al. 1997).

"Traditional" Direct Teaching Versus "New Learning"

In the research literature on teaching effectiveness basically two schools of thought have stood out: a more traditional behaviourist orientation of structured, direct teaching and a "constructivist" inspired orientation, more cognitively oriented, with emphasis on metacognition and self-regulated learning. See Chap. 1, for a more detailed exposure. On rare occasions the "success" of both approaches has been pitted against one another, e.g. Van der Werf (2005). In the Netherlands these two orientations are subject of a sometimes heated debate between "innovators" and "traditionalists" among educational practitioners and commentators under the heading of "new learning". So far the results of meta-analyses favoured the more traditional orientation, see, for example the earlier meta-analyses by Fraser et al. (1987), Walberg (1984), Wang et al. (1993) as well as the review by Van der Werf ibid.).

When considering our results, slightly higher effect sizes are found for constructivist-oriented learning strategies (0.14) as compared to structured, direct teaching (0.09) and teacher-orchestrated classroom management (0.10). Given the earlier findings, this is a striking outcome.

From the number of replications that we obtained in the category constructivist teaching (category IV of the six factor classification in Table 8.13) it is evident that during the last decade a fairly large number of studies addressed the effect of learning to learn, general metacognitive strategies, learning to learn subject specific strategies in mathematics, science and language and cognitive activation. The research agenda of research on teaching appears to have been strongly influenced by the constructivist orientation. Variables like peer tutoring and cooperative learning were included in our category IV as well, although they have a longer tradition and were also part of the earlier meta-analyses (those by Fraser and Walberg and their colleagues), and showed even higher effect sizes in those analyses than in ours.

The constructivist orientation has an enormous societal impact as well, as national educational reforms in countries as different as, for example the Netherlands, Ghana and Turkey, have embraced this orientation as one of the pillars of their educational reform in primary and secondary education.

Turning back to our results it would be too strong a statement to conclude that our results show the superiority of the constructivist-oriented teaching strategies over structured, direct teaching approaches, although, in this context, the difference in effect sizes (0.09 versus 0.14) is not trivial. Comparison of effect sizes on the basis of a meta-analysis in which constructivist strategies are investigated in one set of studies and structured teaching strategies in another should be made with some reservation. In only a few studies constructivist and direct teaching approaches were directly compared. D'Agostino (2000), for example found that more teacher-centred approaches worked better for students in grades one and two of elementary school, whereas a more student-centred, advanced skill approach worked better in grade 4. Such kind of interaction effects with kind of educational objectives, student background characteristics and degree of advancement in the school career were referred to in Chap. 1, and would suggest that constructivist-oriented teaching and more traditional direct teaching approaches could be seen as strategies that teachers might effectively apply in a differentiated way. In other words it might be more fruitful to see these approaches as complementing one another, depending for their effectiveness on aspects of the teaching situation, than to approach those as totally competing strategies (cf. Boekaerts and Simons 1993; Brophy 2001).

The Most Effective Teaching Factors

The factors that had the highest coefficients in the 15-category classification were the following:

An orderly and functional learning environment (0.13); an emotionally supportive climate appeared to be the overriding characterization of this factor. It was striking that variables like achievement pressure and performance orientation had a negative association with achievement. Next, the supportive element in school and classroom climate came out stronger (0.13) in our results than disciplinary climate and orderly classroom management (0.08).

Clear and structured teaching (0.13), including variables as structured, direct teaching and a clear goal-directed teaching approach.

A teaching approach considered "*activating*" (0.12), by means of the application of a range of teaching and learning activities, including authentic task and learning by means of groups and tutors.

Learning to use learning strategies (0.21) came out as the factor with the highest coefficient. It encompasses cooperative learning strategy training, problem-solving, metacognitive training, scientific inquiry training, thinking aloud training, concept mapping, organizing and structuring methods, language acquisition training, phonemic awareness training, reading strategies, writing strategies and formal learning strategy training. The effectiveness of metacognitive strategies is also underlined in the OECD report on students' approaches to learning, OECD (2003). In this study, based on analyses of the PISA 2000 data set, control strategies by means of which students actively monitor their own learning process appeared to be the most powerful strategy. Ideas on "learning to learn" have been around for almost four decades now. In some countries these ideas were the object of heated debates, in which acquiring substantive knowledge was squarely opposed to the teaching of empty procedural skills. Our results demonstrate that these approaches are actually effective, particularly also when they are geared to subject matter-specific strategies, like, for example concept mapping in reading literacy teaching.

A teaching approach characterized as challenging (0.13); as compared to activating, this factor is about stimulating motivation through cognitive challenge rather than through activating teaching and learning in a more organizational sense.

Teaching Variables Showing the Highest Effect Sizes

The overriding conclusion of the meta-analysis of teaching factors is that effective teaching is a matter of clear structuring *and* challenging presentation *and* a supportive climate *and* metacognitive training. The results indicate that these main orientations to teaching are all important, and that effective teaching is not dependent on a singular strategy or approach. This is also illustrated when we take a look at the underlying variables that were used for both the 6- and the 15-factor categorizations. The variables summarized in Table 8.19 have the highest effect sizes, and a fair number of replications (variables with relatively high effect sizes, but less than about 30 replications were not included in Table 8.19).

Table 8.19 Variables with the highest mean effect sizes

Variable	Mean effect size	Number of replications
No achievement pressure	0.15	29
Goal-directed/clear teaching	0.22	36
Cooperative learning	0.20	49
Peer tutoring	0.22	53
Metacognitive training	0.24	35
Scientific inquiry training	0.20	32
Reading and writing strategies	0.21	34
Cognitive activation	0.18	67
Authentic contexts	0.16	47

The Meaning of the Results with Respect to the Models on Teaching and Learning, Discussed in Chap. 2

On the basis of a review of the research literature on teaching effectiveness in Chap. 2, summary models of teaching and learning were provided. Important elements of these models were: an emphasis on contextual and ecological variables, the recognition of "partly malleable dispositions", such as cognitive strategies and engagement, a distinction in proactive, interactive and retroactive malleable teaching variables, and a discussion on structure and independence. This discussion reflects the issue of direct versus constructivist-oriented teaching strategies considered in the above. Our results are more explicit about the cognitive challenge, the learning to learn learning strategies, authentic context and social aspects of constructivism, than about "independence" and self-regulated learning. For the latter we did not find much support, even a (very small) negative effect for individual work (−0.01), based on 39 replications.

Next, our results contained little evidence on the contextual, ecological variables. This latter result was rather due to the effect that these variables were not strongly represented in the articles that were the basis of the analysis, and not a matter of low effect sizes.

The results supported the idea of the relevance of partly malleable cognitive and motivational dispositions.

With respect to the distinction between proactive, interactive and retroactive strategies, most variables addressed phenomena that are basically addressing teacher/learner and learner/learner interactions. The proactive categories regarding time and opportunity to learn indicated effect sizes in the order of 0.12, which, comparatively speaking, should not be seen as small effects. Retroactive strategies concerning evaluation, monitoring and feedback showed effect sizes in the order of 0.07 (see Table 8.13, addressing the six-category classification). Despite the theoretically high potential of these strategies, in actual practice they appear to have less spectacular results, perhaps due to implementation problems (cf. Schildkamp 2007).

Implications of the Results for Educational Policy and Practice

The message of our findings to educational policy and practice would discourage strong antagonism between certain teaching orientations coined in already tendentious terms like "new learning" versus "traditional teaching". Our results do support a place for parts of the constructivist orientation, particularly as far as teaching metacognitive strategies is concerned. Yet, at the same time orderly, teacher-managed classrooms and a clear and structured teaching approach appear to work as well. Outcomes as ours would appear to have more meaning for teacher training and teacher professional development, than as a source of inspiration for basing system-level educational reform on one teaching orientation or the other. In the context of teacher training and professional development it would be important to provide teachers with broad repertoires of teaching strategies that contain elements of direct teaching as well as aspects of a more constructivist-oriented approach.

Suggestions for Research on School and Teaching Effectiveness

The methodological limitations of meta-analyses as described in this chapter have already been stated in previous sections. One obvious limitation is that the analyses treat variables one by one, and do not indicate comparative effect sizes, when different variables are included in one study. Another difficulty is the relative arbitrariness of classifying the basic material. To counter this problem, we have provided considerable attention to the definition of the core concepts at different levels of generality, up to the level of exemplarily items (Chaps. 3 and 5).

A piece of good news from our results was the relatively unimportant impact of most of the moderator variables. The fact that effect sizes in teaching effectiveness studies do not significantly depend on designs being experimental or non-experimental speaks for the robustness of the findings.

Future meta-analyses in this field could benefit from better documentation of studies in databanks, including banks of instruments used to measure the key variables. Perhaps in the future the basic work for meta-analyses could be prepared by documentary services and libraries. In their turn such documentation might help in creating convergence in the definition and operationalization of key variables in this field.

To the degree that a meta-analysis can give rise to suggestions about improving results of the basic research studies in these fields, two issues come to mind. The field of school effectiveness research could benefit from better-controlled longitudinal studies, in which not only the outcome variables are measured repeatedly, but the school characteristics as well. The fragmented results that one obtains from meta-analyses, makes one think of a more ideal situation, particularly in the field of teaching effectiveness, where more complex patterns and sequences of teaching

strategies can be taken as the object of study. In a methodological sense this could be accomplished by means of direct observation and video registration. To the degree that results of reviews like the one discussed in this report as well as conceptual models on effective teaching reflect a growing consensus on the kind of factors that are to be studied, such observation studies might become less inductive and more theory or model driven.

Annex 1: Steps in the Rechecking of the Database from Scheerens et al. (2005)

The Way in Which Non-significant Effects Are Incorporated[3]

Some authors present only values for significant effects in the published tables, whereas the non-significant effects are left blank. This can especially be the case in multi-level and regression analyses. For these publications originally only the values for the reported (and thus significant) variables were entered in the database for meta-analyses. In the correction stage the effects which were left blank were added to the database with effect zero.

Standardized and Non-standardized Effects

For effects based on multi-level and regression analyses it is a major issue, whether or not the reported effects are standardized. For meta-analyses the effects should be standardized. It turned out that for some publications originally a wrong interpretation had been used (i.e. standardized effects, where they were not standardized, or vice versa). In the correction stage many effects were adapted for this reason.

In some cases the adjusted effects became considerably lower, in others considerably higher.

Converting Effect Size Indicators

For some effect measures, like F-tests with d$f > 1$, it is impossible to convert the measures to an effect estimate which can be used in meta-analyses. Depending on the degree of contextual information given in the original publication, in the correction stage these cases have been considered as either missing or an effect value was assigned, which reflected the reported significance level.

[3]Technical analyses for this chapter were authored by Rien Steen.

Correcting Various Kinds of Errors in the Original Tabulations

For a few cases the sign of the effect has been reversed in the correction stage.

In some instances the number of cases has been adapted, which leads to a different standardized effect where standardization on the original effects is applied.

Finally, in a few cases the results were originally misinterpreted. Adaptations depended on the contextual information provided in the original publication.

Combination of Effects

Variables in the original studies are considered to be indicators for broader concepts, called sub-areas in this meta-analysis. It turned out that in some instances two or more of the original variables in a multi-level analysis were considered to be indicators for the same concept. In the present analysis, in such cases the average effect of these variables were counted as one effect instead of the separate effects used in the present meta-analysis This has only been done for variables involved in the same analysis, like a multi-level analysis or a regression analysis. Where separate analyses were reported, like simple correlations or t-tests, effects have not been averaged.

Where broader categorization is used, i.e. the categorization using 15 or 6 categories, further averaging of original effects was performed, where appropriate. This also implies that the numbers of replications reported for the broader categories may be less than the sum of the numbers of replications reported for the basic categories.

Annex 2: Mean Effect Sizes and 95 % Confidence Intervals for the 15 and 6 Factor Categorization of the Teaching Variables

Area		Mean eff.	St. uerror	Lower	Upper	Count
1	Learning time	0.095	0.013	0.069	0.121	173
2	Classroom organization	0.075	0.019	0.038	0.111	62
3	Learning environment	0.129	0.012	0.105	0.153	138
4	Clear and structured	0.126	0.016	0.095	0.157	134
5	Activating	0.123	0.012	0.098	0.147	179
6	Learning strategies	0.213	0.018	0.177	0.248	103
7	Challenge	0.130	0.013	0.105	0.155	180

(continued)

(continued)

Area		Mean eff.	St. uerror	Lower	Upper	Count
8	Support	0.108	0.017	0.075	0.140	73
9	Feedback	0.056	0.019	0.018	0.094	106
10	Evaluation	0.086	0.031	0.025	0.148	46
11	Teacher characteristics	0.146	0.039	0.070	0.222	26
12	Adaptive teaching	0.066	0.027	0.014	0.118	41
13	Practice	−0.080	0.030	−0.138	−0.021	27
14	Material	0.015	0.016	−0.015	0.046	28
15	Integrative approaches	0.089	0.022	0.046	0.131	90

Category		Mean eff.	St. error	Lower	Upper	Count
I	Curricular	0.077	0.023	0.032	0.122	61
II	Teacher-orchestrated classroom management	0.095	0.010	0.076	0.114	304
III	Teaching strategy (structured, direct, mastery)	0.087	0.015	0.057	0.116	165
IV	Teaching strategy (constructivist-oriented ….)	0.135	0.008	0.120	0.150	542
V	Climate, support, positive interactions	0.117	0.011	0.096	0.138	180
VI	Feedback/monitoring/assessment/tests	0.065	0.017	0.033	0.098	152

Annex 3: Technical Annex: Statistical Modelling Issues

The multi-level model for the school-level meta-analysis is:

$$\begin{aligned}\delta_{rs} = {} & \delta_0 + \gamma_1\,\text{subject-math}_{rs} + \gamma_2\,\text{subject-lang}_{rs} + \gamma_3\,\text{sector}_s \\ & + \gamma_4\,\text{country-USA}_s + \gamma_5\,\text{country-NL} + \gamma_6\,\text{design}_s \\ & + \gamma_7\,\text{statistical technique employed}_s + \gamma_8\,\text{level}_s + u_{rs} + v_s + e_{rs}(\cdot)\end{aligned}$$

where

d_{rs} is the effect size d in replication r in study s, which is an estimate of the population parameter δ_{rs}

e_{rs} is the associated sampling error with δ_{rs}

δ_0 is the effect size across studies

v_s is the associated sampling error with δ_0 (the across replications sampling error)

u_{rs} is the associated sampling error with δ_s (the within replications sampling error)

γ_1 through γ_8 are coefficients with the following predictors:

Subject-math	0 = not math only, 1 = math only
Subject-lang	0 = not language only, 1 = language only
Sector	0 = primary education, 1 = secondary education
Country-USA	0 = not USA, 1 = USA
Country-NL	0 = not The Netherlands, 1 = The Netherlands
Design	0 = gross, 1 = value added (correction for prior achievement and/or background variables)
Statistical technique employed	0 = multi-level technique, 1 = not multi-level technique
Level	0 = teacher/class level, 1 = school/school leader level

The multi-level for the student level differs with respect to the country moderator variable and the research design moderator (experimental or correlational).

$$\delta_{rs} = \delta_0 + \gamma_1\,\text{subject-math}_{rs} + \gamma_2\,\text{subject-lang}_{rs} + \gamma_3\,\text{sector}_s + \gamma_4\,\text{country-USA}_s + \gamma_5\,\text{country-NL} + \gamma_6\,\text{country-GER} + \gamma_7\,\text{design}_s + \gamma_8\,\text{statistical technique employed}_s + \gamma_9\,\text{experimental} + u_{rs} + v_s + e_{rs}(\cdot)$$

where

d_{rs} is the effect size d in replication r in study s, which is an estimate of the population parameter δ_{rs}

e_{rs} is the associated sampling error with δ_{rs}

δ_0 is the effect size across studies

v_s is the associated sampling error with δ_0 (the across replications sampling error)

u_{rs} is the associated sampling error with δ_s (the within replications sampling error)

γ_1 through γ_9 are coefficients with the following predictors:

Subject-math	0 = not math only, 1 = math only
Subject-lang	0 = not language only, 1 = language only
Sector	0 = primary education, 1 = secondary education
Country-USA	0 = not USA, 1 = USA
Country-NL	0 = not The Netherlands, 1 = The Netherlands
Country-GER	0 = not Germany, 1 = Germany
Design	0 = gross, 1 = value added (correction for prior achievement and/or background variables)
Statistical technique employed	0 = multi-level technique, 1 = not multi-level technique
Experimental	0 = correlational, 1 = (Quasi-) experimental

Thus in the equation δ_0 is the estimated effect size for studies where all predictors have value 0.

The model is the same as used by Bosker and Witziers (1995), which is based on a model from Raudenbush (cf. Raudenbush and Bryk 1986; Raudenbush 1994).

Note:

Moderator *subject* consists of three categories, i.e. "math only", "language only" and a rest category, mainly containing composite scores and subject science. Because only binary variables can be handled in the above equation, for subject-math, category 0 contains both "language" and the rest category. Likewise for subject-lang, category 0 contains both "math" and the rest category.

For the analysis of the school factors, the moderator *country* consists of three categories, i.e. "USA", "The Netherlands" and "countries not being USA or The Netherlands". Where in the text a comparison is made between the USA and "other countries", these other countries include the Netherlands. Likewise, where the Netherlands is compared with "other countries", these other countries include the USA.

For the analysis of the teaching factors the moderator *country* consists of four categories, i.e. "USA", "The Netherlands", "Germany" and "countries not being USA, the Netherlands or Germany". Where in the text a comparison is made between the USA and "other countries", these other countries include the Netherlands and Germany. The same principle applies where the Netherlands or Germany are compared with "other countries".

Effect Sizes

In the studies from which the results are analysed here, the effect sizes have been reported in various ways. Most of the effect sizes could be transformed directly into Fisher's Z using formulae presented by Rosenthal (1994).

The most used effect sizes are regression coefficients from multiple regression or multi-level analyses, t-tests, correlations and path analyses coefficients.

In a number of studies however, only information about the significance of effects has been reported in various ways. The most common ways were: significant/not significant at a given significance level, two or three significance levels for more significant effects and a p-value for an effect. In those cases conservative estimates of the effects have been used in the meta-analysis, i.e. the effects have been estimated as being just significant for the reported significance level, which means that the absolute values will have been underestimated. Where effects were only reported as being not significant, the effect has been estimated as zero. This is also the case, usually in ML analyses, where no values were reported if the effect was not significant.

In many studies results have been reported in various ways and/or for different models and/or using various techniques. Where a relation between an indicator of

concepts relating to teacher and/or school effectiveness and an outcome indicator has been reported for different models, only the result of the most complex model has been used in the meta-analysis.

Where in one original analysis (e.g. a ML analysis) two or more indicators of the same concept were used, the average of the Fisher's Z values was used in the meta-analyses.

Weighting

The weights used in the meta-analyses are based on a random effects model. In this model the relative weights depend on both the sample sizes N_i as used in each original analysis and on the variance of the original effect measures.

The weights are computed as

$$\begin{aligned}\text{weight}_i &= 1/\text{var}\,(\text{estimate of the effect size } T_i)\\ &= 1/(\text{var}_i(\text{Fishers } Z) + \text{var}\,(\text{Fishers } Z_i)),\ (\text{cf. Raudenbush 1994, formula 20.3})\end{aligned}$$

where $\text{var}_i(\text{Fishers } Z) = 1/(N_i - 3)$.

So the first variance depends on the sample size N_i of a study, the second variance is an overall variance over the estimates of Fisher's Z in all studies used in each meta-analysis.

In order to both reduce the chance factor and to make the computations more simple, var (Fishers Z_i) is based on all studies in all analyses and is found to be 0.041. So

$$\text{weight}_i = 1/(1/(N_i - 3) + 0.041).$$

References[4]

School Effectiveness Part

Anderson, C. S. (1982a). The search for school climate: A review of the research. *Review of Educational Research, 52*(3), 368–420.

Baumert, J., Lüdtke, O., & Trautwein, U. (2006). Interpreting effect sizes in large-scale educational assessments. Internal Paper Max Planck Institute for Human Development, Berlin.

Borman, K. M., Carter, K., Aladjem, D. K., & LeFloch, K. C. (2004). Challenges for the future of comprehensive school reform. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from*

[4]References in the list not referred to in the text of the chapter refer to studies used in the meta-analyses.

comprehensive school reform research (p. 150). Washington, DC: The National Clearinghouse for Comprehensive School Reform.

Bosker, R. J., & Scheerens, J. (1995). A self-evaluation procedure for schools using multilevel modelling. *Tijdschrift voor Onderwijsresearch, 20*(2), 154–164.

Bosker, R. J., de Vos, H., & Witziers, B. (2000). *Theories and models of educational effectiveness*. Enschede: University of Twente, Department of Educational Organization and Management.

Bosker, R. J., & Witziers, B. (1995). *A meta analytical synthesis of school effectiveness research: The size of school effects and the effect size of educational leadership*. Enschede: University of Twente, Department of Educational Organization and Management.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 193–214). New York: Russell Sage Foundation.

Carnoy, M., Elmore, R., & Siskin, L. (2003). (Eds.). *The new accountability. high schools and high-stakes testing*. New York/London: Routledge Falmer.

Cohen, M. (1982, January–February 13–16). Effective schools: Accumulating research findings. *American Education*.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cotton, K. (1995). *Effective schooling practices: A research synthesis*. 1995 Update. School Improvement Research Series. Northwest Regional Educational Laboratory.

Creemers, B. P. M. (1994a). *The effective classroom*. London: Cassell.

Creemers, B., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366.

Dougherty, K. (1981a). After the fall: Research on school effects since the Coleman report. *Harvard Educational Review, 51*, 301–308.

Edmonds, R. R. (1979a). Some schools work and more and more can. *Social Policy, 9*, 28–32.

Fuller, B. (1987). What factors raise achievement in the Third World? *Review of Educational Research, 57*, 255–292.

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules and pedagogy. *Review of Educational Research, 64*, 119–157.

Good, T. L., & Brophy, J. (1986). School effects. In: M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: McMillan Inc.

Hallinger, Ph, & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement, 9*(2), 157–191.

Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *The Elementary School Journal, 96*(5), 527–550.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *The World Bank Research Observer, 10*, 227–246.

Hill, P. W., Rowe, K. J., & Holmes-Smith, P. (1995). *Factors affecting students' educational progress: Multilevel modelling of educational effectiveness*. Paper Presented at the 8th International Congress for School Effectiveness and Improvement, Leeuwarden, the Netherlands, January 3–6, 1995.

Hofman, A. W. H. (1995). Cross-level relationships within effective schools. *School Effectiveness and School Improvement, 6*, 146–174.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kyle, M. J. (Ed.). (1985). *Reaching for excellence. An effective schools sourcebook*. Washington, DC: U.S. Government Printing Office.

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement, 16*(2), 103–152.

Kythreotis, A., & Pashiardis, P. (2006). The influences of school leadership styles and culture on students' achievement in Cyprus primary schools. *AERA* (San Francisco).

Levine, D. K., & Lezotte, L. W. (1990a). *Unusually effective schools: A review and analysis of research and practice*. Madison, Wise: Nat. Centre for Effective Schools Research and Development.

Lockheed, M., & Hanushek, E. (1988). Improving educational efficiency in developing countries: What do we know? *Compare, 18*(1), 21–38.

Lockheed, M., & Verspoor, A. (1991). *Improving primary education in developing countries*. London: Oxford University Press.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD.

Miles, M. (1998a). finding keys to school change: A 40-year Odyssey. In A. Hargreaves, A. Liebermann, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change*. Dordrecht/Boston/London: Kluwer Publishing Company.

Muijs, D., & Reynolds, D. (2001a). *Effective teaching*. London: Paul Chapman.

Murnane, R. J. (1981a). Interpreting the evidence on school effectiveness. *Teachers College Record, 83*, 19–35.

Neufeld, E., Farrar, E., & Miles, M. B. (1983a). *A review of effective schools research: The message for secondary schools*. Washington, DC: National Commission on Excellence in Education.

Purkey, S. C., & Smith, M. S. (1983a). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Ralph, J. H., & Fennessey, J. (1983a). Science or reform: Some questions about the effective schools model. *Phi Delta Kappan, 64*(10), 689–695.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Russell Sage Foundation.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1–17.

Reezigt, G. J., Guldemond, H., & Creemers, B. P. M. (1999). Empirical validity for a comprehensive model on educational effectiveness. *School Effectiveness and School Improvement, 10*(2), 193–216.

Reynolds, D., Hopkins, D., & Stoll, L. (1993a). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement, 4*(1), 37–58.

Richard, F. D., Bond, C. F, Jr, & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331–363.

Rosenthal, R. (1994a). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Ross, S. M., & Gil, L. (2004). The past and future of comprehensive school reform: Perspectives from a researcher and practitioner. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 151–174). Washington, DC: The National Clearinghouse for Comprehensive School Reform.

Rowan, B., Camburn, E., & Barnes, C. (2004). Benefiting for comprehensive school reform: A review of research on CSR implementation. In C. T. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 1–52). Washington, DC: The National Clearinghouse for Comprehensive School Reform.

Rutter, M. (1983a). School effects on pupil progress: Research findings and policy implications. *Child Development, 54*(1), 1–29.

Sammons, P., Hillman, J., & Mortimore, P. (1995a). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J., & Bosker, R. J. (1997a). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Schildkamp, K. (2007). *The utilization of a self-evaluation instrument for primary education*. Enschede: University of Twente, Department of Educational Organization and Management (dissertation).

Seidel, T., & Steen, R. (2005). The indicators on teachin and learning compared to the review of recent research articles on school and instructional effectiveness. In: J. Scheerens, T. Seidel, B. Witziers, M. Hendriks, & G. Doornekamp. (2005). *Positioning the supervision frameworks for primary and secondary education of the Dutch Educational Inspectorate in current educational discourse and validating core indicators against the knowledge base of educational effectiveness research*. Enschede/Kiel: University of Twente/Institute for Science Education (IPN).

Slavin, R. E. (2000). *Educational psychology: Theory and practice* (6th ed.). Needham Heights, MA: Allyn and Bacon.

Stoll, L., & Myers, K. (1997). *No quick fixes. Perspectives on schools in difficulty*. London: Falmer Press.

Stringfield, S. (1994). Outlier studies of school effectiveness. In D. Reynolds, et al. (Eds.), *Advances in school effectiveness research and practice*. Oxford: Pergamon.

Stringfield, S.C., & Slavin, R.E. (1992). A hierarchical longitudinal model for elementary school effects. In B. P. M. Creemers & G. J. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35–39). Groningen: ICO.

Sweeney, J. (1982a). Research synthesis on effective school leadership. *Educational Leadership, 39*, 346–352.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.

Tymms, P., Merrell, C., & Henderson, B. (1997). The first year at school. A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation, 3* (2), 101–118.

van de Grift, W. J. C. M., & Houtveen, A. A. M. (2006). Underperformance in primary schools. *School Effectiveness and School Improvement, 17*(3), 255–274.

Walberg, H. J. (1984). Improving the productivity of American schools. *Educational Leadership, 41*, 19–27.

Witziers, B., Bosker, R. J., & Krüger, M. (2003a). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39*(3), 398–425.

Teaching Effectiveness Part

Ackoff, R. L. (1981). *Creating the corporate future*. New York: Wiley.

Anderman, E. M., Eccles, J. S., Yoon, K. S., Roeser, R., Wigfield, A., & Blumenfeld, P. (2001). Learning to value mathematics and reading: Relations to mastery and performance-oriented instructional practices. *Contemporary Educational Psychology, 26*(1), 76–95.

Anderson, C. S. (1982b). The search for school climate: A review of the research. *Review of Educational Research, 52*(3), 368–420.

Anderson, L. W. (1991). *Increasing teacher effectiveness*. Paris: UNESCO, IIEP.

Anderson, L. W. (2004). *Increasing teacher effectiveness* (2nd ed.). Paris: UNESCO, IIEP.

Anderson, L. W., & Pellicer, L. O. (1998). Towards an understanding of unusually successful programs for economically disadvantaged students. *Journal of Education for Students Placed at Risk, 3*, 237–263.

Arnold, D. H., Fisher, P. H., Doctoroff, G. L., & Dobbs, J. (2002). Accelerating math development in head start classrooms. *Journal of Educational Psychology, 94*(4), 762–770.

Artelt, C., Baumert, J., McElvany, N. J., & Peschar, J. (2003). *Learners for life. Student approaches to learning. Results from PISA 2000*. Paris: OECD.

Bangert, R. L., Kulik, J. A., & Kulik, C. C. (1983). Individualized systems of instruction in secondary schools. *Review of Educational Research, 53*, 143–158.

Baumert, J., Blum, W., Brunner, M., Jordan, A., Klusmann, U., Krauss, S., et al. (2005). *Professional knowledge of teachers, cognitively activating instruction, and the development of*

mathematical competence (COACTIV). Presentation at the Meeting of the OECD INES Taskforce on Teaching Effectiveness, Reykjavik, 2005.

Baumert, J., Blum, W., & Neubrand, M. (2001). *Surveying the instructional conditions and domain-specific individual prerequisites for the development of mathematical competencies*. Paper Presented at the Special Meeting on the Co-ordination of National Teacher Components for PISA 2003, Munich, Germany, May 28, 2001.

Baumert, J., Blum, W., & Neubrand, J. (2002). *Drawing lessons from PISA 2000—Long-term research implications: Gaining a better understanding of the relationship between system inputs and learning outcomes by assessing instructional and learning processes as mediating factors*. Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, J., Carstensen, C. H., & Siegle, T. (2005). Wirtschaftliche, soziale und kulturelle Lebensverhältnisse un regionale Disparitäten des Kompetenzerwerbs. In PISA-Konsortium Deustschland (Hrsg.) *PISA 2003. Der zweite Vergleich der Länder Deutschlands – Was wissen und können Jugendliche?* Münster/New York/Berlin/München, 2005.

Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Eds.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematisch und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2* (pp. 271–315). Opladen: Leske+Budrich.

Bell, P. (1998). *Designing for students' science learning using argumentation and classroom debate*. Berkeley, CA: University of California.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Bloom, B. S., Hastings, J Th, & Madaus, G. F. (1971). *Handbook on formative and summative evalaution of student learning*. New-York: MacGraw-Hill.

Boekaerts, M., & Simons, P. R. J. (1993). *Leren en Instructie. Psychologie van de Leerling en het Leerproces*. Assen: Dekker & Van de Vegt.

Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. *Learning and Instruction, 13*(3), 327–347.

Borich, G. D., & Jemelka, R. P. (1982). *Programs and systems. An evaluation perspective*. New York: Academic Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brophy, J. (2001). Generic aspects of effective teaching. In M. C. Wang & H. J. Walberg (Eds.), *Tomorrow's teachers*. Berkeley: McCutchan Publishing Company.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Third handbook of research on teaching* (pp. 328–375). New York: MacMillan.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1997).

Byrne, B., & Fieldingbarnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young-children—A 2-year and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*(3), 488–503.

Cantrell, S. C. (1999). The effects of literacy instruction on primary students' reading and writing achievement. *Reading Research and Instruction, 39*(1), 3–26.

Carbonell, J. R. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems, 11*(4), 190–202.

Carroll, J. B. (1963). A model of school learning. *Teachers college record, 64*, 722–733.

Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*(3), 248–263.

Clausen, M. (2001). *Unterrichtsqualität: eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.

Clauset, K. H., & Gaynor, A. K. (1982). A systems perspective on effective schools. *Educational Leadership, 40*(3), 54–59.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
Cohen, M. (1982, January-February 13–16). Effective schools: Accumulating research findings. *American Education*.
Cohen, D. K. (1988). Teaching practice … Plus ça change … In P. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice*. Berkeley, CA: McCutchan.
Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction* (pp. 453–495). Hillsdale: Lawrence Erlbaum.
Collins, A., & Stevens, A. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. II). Hillsdale, NJ: Lawrence Erlbaum Associates.
Cotton, K. (1995). *Effective schooling practices: A research synthesis*. 1995 Update. School Improvement Research Series. Northwest Regional Educational Laboratory.
Creemers, B. P. M. (1994b). *The effective classroom*. London: Cassell.
Cronbach, L. J., & Snow, R. E. (1981). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington Publishers.
Cunningham, D. J. (1991). In defense of extremism. *Educational Technology, 31*(9), 26–27.
D'Agostino, J. V. (2000). Instructional and school effects on students' longitudinal achievement in reading and mathematics achievement. *School Effectiveness and School Improvement., 11*(2), 197–235.
Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1). Available: http://olam.ed.asu.edu/epaa/v8n1/
Davies, J. K. (1972). Style and effectiveness in education and training: A model for organizing, teaching and learning. *Instructional Science*, *2*.
De Corte, E., et al. (1973). *Beknopte didaxologie*. Groningen: H.D. Tjeenk Willink.
de Groot, A. D., & van Naerssen, H. (1966). *Vijven en zessen*. Wolters: Cijfers en beslissingen; het selectieproces in ons onderwijs. Groningen.
de Jong, T. (2006). Computer simulations: Technological advances in inquiry learning. *Science, 312*(5773), 532–533.
de Klerk, L. F. W. (1985). ATI-onderzoek en differentiatie: een reactie. *Pedagogische Studiën, 62*, 372–375.
de Leeuw, A. C. J. (1990). *Organisaties: Management, analyse, ontwerp en verandering: een systeemvisie*. Assen/Maastricht: Van Gorcum.
de Vos, H. (1989). A rational-choice explanation of composition effects in educational research. *Rationality and Society, 1*(2), 220–239.
Dougherty, K. (1981b). After the fall: Research on school effects since the Coleman report. *Harvard Educational Review, 51*, 301–308.
Doyle, W. (1985). Effective secondary classroom practices. In: M. J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook*. Washington, DC: US Government Printing Office.
Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: MacMillan.
Duffy, Th M, & Jonassen, D. H. (1992). *Constructivism and the technology of instruction: A conversation*. Hillsdale, NJ: Lawrence Erlbaum Ass.
Edmonds, R. R. (1979b). *A discussion of the literature and issues related to effective schooling*. Cambridge: Center for Urban Studies, Harvard Graduate School of Education.
Einsiedler, W., & Treinies, G. (1997). Effects of teaching methods, class effects, and patterns of cognitive teacher-pupil interactions in an experimental study in primary school classes. *School Effectiveness and School Improvement, 8*(3), 327–353.
Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *Special Issue of the International Journal of Educational Research*, *11*(2).
Gage, N. (1965). Desirable behaviors of teachers. *Urban Education, 1*, 85–95.
Gagne, R. M. (1972). *Conditions of learning*. New York: Holt, Rinehart and Winston.

Good, T. L., & Brophy, J. (1986). School effects. In: M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: McMillan Inc.

Gruehn, S. (1995). The compatibility of cognitive and noncognitive objectives of instruction. *Zeitschrift Für Padagogik, 41*(4), 531–553.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14*, 351–388.

Hanushek, E. A. (1989). Expenditures, efficiency, and equity in education: The federal Government's role. *The American Economic Review, 79*(2), 46.

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *The World Bank Research Observer*, *10*, 227–246.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*, 141–164.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Hill, P. W., & Rowe, K. J. (1998). Modelling student progress in studies of educational effectiveness. *School Effectiveness and School Improvement, 9*(3), 310–333.

Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching, 36*(10), 1085–1109.

Irwin, C. C. (1986). What research tells the principal about educational leadership. *Scientica Paedagogica Experimentalis, 23*, 124–137.

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research and Development, 39*(3), 5–14.

Jonassen, D. H. (1992). Evaluating constructivist learning. In Th M Duffy & D. H. Jonassen (Eds.), *Constructivism and the Technology of instruction: A conversation* (pp. 138–148). Hillsdale, NJ: Lawrence Erlbaum Ass.

Kerr, S. (1977). Substitutes for leadership: Some implications for organizational design. *Organizational and Administrative Sciences, 8*, 135–146.

Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000 - Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 333–360). Opladen: Leske+Budrich.

Klieme, E., & Reusser, K. (2003). Unterrichtsqualität und mathematisches Verständnis im internationalen Vergleich. Ein Forschungsprojekt und erste Schritte zur Realisierung. *Unterrichtswissenschaft: Forschung und Entwicklung, Programmierung, Technologie*, *31*(3), 194–205.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin, 119*(2), 254–284.

Kobarg, M. (2004). *Die Bedeutung prozessorientierter Lernbegleitung für kognitive und motivationale Prozesse im Physikunterricht – eine Videostudie*. IPN / Fachbereich Psychologie der CAU Kiel: Unveröffentlichte Diplomarbeit. Kiel.

Kulik, C. L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of research findings. *American Educational Research Journal, 19*, 415–428.

Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.

Kyle, M. J. (Ed.). (1985). *Reaching for Excellence. An effective schools sourcebook*. Washington, DC.: U.S. Government Printing Office.

Laarhoven, P. van, & Vries, A. M. de. (1987). Effecten van de interklassikale groeperingsvorm in het voortgezet onderwijs. Resultaten van een literatuurstudie. (Effects of heterogeneous grouping in secondary schools). In: J. Scheerens & W.G.R. Stoel (Red.), *Effectiviteit van onderwijsorganisaties*. Lisse: Swets & Zeitlinger.

Leithwood, K. A., & Montgomery, D. J. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research, 52*, 309–399.

Levine, D. K., & Lezotte, L. W. (1990b). *Unusually effective schools: A review and analysis of research and practice*. Madison, Wise: Nat. Centre for Effective Schools Research and Development.

Linn, M. C., et al. (2006). Inquiry learning: Teaching and assessing knowledge integration in science. *Science, 313*(5970), 1049–1050.

Lortie, D. C. (1973). Observations on teaching as work. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B. U., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD.

McNeill, J. D. (1969). Forces influencing curriculum. *Review of Educational Research, 39*(3), 293–318.

Medley, D., & Mitzel, H. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Merrill, M. D. (1991). Constructivism and instruction design. *Educational Technology, 31*, 45–53.

Miles, M. (1998b). Finding keys to school change: A 40-year Odyssey. In A. Hargreaves, A. Liebermann, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change*. Dordrecht/Boston/London: Kluwer Publishing Company.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and Analysis in task-based language assessment. *Language Testing, 19*(4), 477–496.

Mitchell, D. E., & Tucker, Sh. (1992). Leadership as a way of thinking. *Educational Leadership, 49*(5), 30.

Monk, D. H. (1989). The education production function: It's evolving role in policy analysis. *Educational Evaluation and Policy Analysis, 11*(1), 31–45.

Monk, D. H. (1992). *Microeconomics of School Productions*. Paper for the Economics of Education Section of the International Encyclopedia of Education.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset: Open books.

Muijs, D., & Reynolds, D. (2001b). *Effective teaching*. London: Paul Chapman.

Murnane, R. J. (1981b). Interpreting the evidence on school effectiveness. *Teachers College Record, 83*, 19–35.

National Center for Education Statistics (NCS). (2002). *Fundamental problems in the measurement of instructional processes*. Washington, DC: NCS.

Neufeld, E., Farrar, E., & Miles, M. B. (1983b). *A review of effective schools research: The message for secondary schools*. Washington, DC: National Commission on Excellence in Education.

Nolen, S. B. (2003). Learning environment, motivation, and achievement in high school science. *Journal of Research in Science Teaching, 40*(4), 347–368.

OECD. (2003). *Learners for life. Student approaches to learning. Results from PISA 2000.* Paris: OECD.

Opdenakker, M., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*, 407–432.

Postlethwaite, T. N., & Ross, K. N. (1992). *Effective schools in reading. implications for educational planners*. The Hague: IEA.

Prenzel, M., & Ostermeier, C. (in press). What can we learn from different forms of evaluation: Experiences of quality development program in mathematics and science instruction. In J. Bennett, J. Holeman, & R. Millar (Hrsg.), *Making the difference: Evaluation as a tool for improving science education*. Münster: Waxmann.

Purkey, S. C., & Smith, M. S. (1983b). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Ralph, J. H., & Fennessey, J. (1983b). Science or reform: Some questions about the effective schools model. *Phi Delta Kappan, 64*(10), 689–695.

Ravitz, J. L., Becker, H. J., & Wong, Y. (2000). *Teaching, learning, and computing: 1998 national survey*. University of California, Irvine: Center for Research on Information Technology and Organizations and University of Minnesota.

Raudenbush, S. (1994b). Searching for balance between a priori and post hoc model modification: Is a 'general approach' desirable? *School effectiveness and school improvement, 5*(2), 196–198.

Reezigt, G. J. (1993). *Effecten van differentiatie op de basisschool (Effects of grouping in primary education)*. Groningen: RION.

Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academic Press.

Reynolds, D., Hopkins, D., & Stoll, L. (1993b). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement, 4*(1), 37–58.

Rosenshine, B. (1979). Content, time, and direct instruction. In P. Peterson & H. J. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications*. Berkeley: McCutchan.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.

Rosenthal, R. (1994b). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166–169.

Rutter, M. (1983b). School effects on pupil progress: research findings and policy implications. *Child Development, 54*(1), 1–29.

Sammons, P., Hillman, J., & Mortimore, P. (1995b). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J. (1987). *Enhancing educational opportunities for disadvantaged learners*. Amsterdam: North-Holland Publishing Comp.

Scheerens, J. (1989). Process indicators of school functioning. *School Effectiveness and School Improvement, 1*(1), 61–80.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, *1*(1), 61–80 (Lisse: Swets & Zeitlinger).

Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.

Scheerens, J. (1994). The school-level context of instructional effectiveness: A comparison between school effectiveness and restructuring models. *Tijdschrift voor Onderwijsresearch, 19* (1), 26–38.

Scheerens, J. (2000). *Improving school effectiveness* (p. 68). Paris: UNESCO, IIEP, Fundamentals of Educational Planning series no.

Scheerens, J., & Bosker, R. J. (1997b). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. In: B. P. M. Creemers & J. Scheerens (Eds.), *Developments in school effectiveness research* (pp. 691–706). Special issue of *the International Journal of Educational Research*, (13)7. Oxford: Pergamon Press.

Scheerens, J., Vermeulen, C. J. A. J., & Pelgrum, W. J. (1989). Generalizability of school and instructional effectiveness indicators across nations. In: B. P. M. Creemers & J. Scheerens (Eds.), *Development in school effectiveness research*. Oxford: Pergamon Press (Special issue of the *International Journal of Educational Research*, *13*(7), 789–800).

Scheerens, J., Seidel, T., Witziers, B., Hendriks, M., & Doornekamp, G. (2005). *Positioning the supervision frameworks for primary and secondary education of the Dutch Educational Inspectorate in current educational discourse and validating core indicators against the knowledge base of educational effectiveness research*. Enschede/Kiel: University of Twente/Institute for Science Education (IPN).

Schulmeiß, I. (2004). *Vermischung von Lern- und Leistungssituationen und ihre Bedeutung für die Qualität der Lernmotivation - eine Videostudie im Physikunterricht*. IPN, Institut für Psychologie der CAU Kiel (unveröffentlichte Diplomarbeit): Kiel.

Seidel, T. (2003). *Lehr-Lernskripts im Unterricht*. Münster: Waxmann.

Seidel, T., & Prenzel, M. (2003). Mit Fehlern umgehen - Zum Lernen motivieren. *Praxis der Naturwissenschaften - Physik in der Schule, 52*(1), 30–34.

Seidel, T., Rimmele, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction, 15*(6), 539–556.

Seidel, T., & Steen, R. (2005). The indicators on teaching and learning compared to the review of recent research articles on school and instructional effectiveness. In J. Scheerens, T. Seidel, B. Witziers, M. Hendriks, & G. Doornekamp (Eds.), *Positioning the supervision frameworks for primary and secondary education of the dutch educational inspectorate in current educational discourse and validating core indicators against the knowledge base of educational effectiveness research*. Enschede/Kiel: University of Twente/Institute for Science Education (IPN).

Slavin, R. E. (1987). *Cooperative learning: Student teams* (2nd ed.). Washington, DC: National Education Association Professional Library.

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston: Allyn & Bacon.

Slavin, R. E. (1996). *Success for All*. Lisse: Swets & Zeitlinger.

Spiro, R. J., Feltowich, P. J., Jacobson, M. J., & Caulson, R. L. (1992). Cognitive flexibility, constructivism and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In Th M Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation*. Hillsdale, NJ: Lawrence Erlbaum Ass.

Stallings, J. (1985). Effective elementary classroom practices. In M. J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook*. Washington, DC: US Government Printing Office.

Stallings, J., & Mohlman, G. (1981). *School policy, leadership style, teacher change and student behavior in eight schools*. Final report to the National Institute of Education, Washington, DC.

Stanat, P. (2003). Predicting Achievement of Immigrant Students: The Role of Student Background at the Level of Individuals and Schools. Presentation at the 10th Biannual EARLI Conference in Padova, Italy, August 26–30, 2003.

Stringfield, S. C., & Slavin, R. E. (1992). A hierarchical longitudinal model for elementary school effects. In B. P. M. Creemers & G. J. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35–39). Groningen: ICO.

Sweeney, J. (1982b). Research synthesis on effective school leadership. *Educational Leadership, 39*, 346–352.

Tobias, S. (1991). An eclectic examination of some issues in the constructivist-ISP controversy. *Educational Technology, 31*(9), 41–43.

van Merriënboer, J. J. G., & Kanselaar, G. (2006). Waar staan we na 25 jaar onderwijstechnologie in Vlaanderen, Nederland en de rest van de wereld? *Pedagogische Studiën, 83*(4), 278–300.

van der Hoeven-Van Doornum, A. A. (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen [The impact of teachers' expectations on educational careers]*. Nijmegen: ITS/OOMO.

van der Werf, M. P. C. (2005). *Leren in het studiehuis: consumeren, construeren of engageren?* Groningen: GION. Rede Universiteit Groningen.

Vedder, P. H. (1985). *Cooperative learning. A study on processes and effects of cooperation between primary school children*. The Hague: SVO.

Veenman, M. V. J., & Hout-Wolters, B. H. A. M. van. (2002). Het meten van metacognitieve vaardigheden. In F. Daems, R. Rymenans, & G. Rogiest (Eds.), *Onderwijsonderzoek in Nederland en Vlaanderen. Proceedings 29e ORD 2002*. (pp. 102–103). Wilrijk: Universiteit van Antwerpen.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*, 249–294.

Weeda, W. C. (1986). Effectiviteitsonderzoek van scholen. In: J. C. van der Wolf & J. J. Hox (Red.), *Kwaliteit van het onderwijs in het geding* (About education quality). Publicaties van het Amsterdams Pedogogische Centrum, nr. 2. Lisse: Swets & Zeitlinger.

Weick, K. E. (1969). *The social psychology of organizing*. Reading, Ma: Addison-Wesley Pub.

Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. *Educational Evaluation and Policy Analysis, 10*(12).

White, B. Y., & Frederikson, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

Willms, J.D. (2004). *Contextual effects on student outcomes*. University of New Brunswick, Canadian Research Institute for Social Policy.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003b). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39* (3), 398–425.

Woesmann, L. (2005). Educational production in Europe. *Economic Policy*, *43*(20), 445–504.

Chapter 9
System-Level Conditions: Miscellaneous Results

Keywords System effectiveness · School autonomy · Private schools · School choice · Horizontal and vertical differentiation of education systems · Loose coupling · Alignment · Accountability · Comprehensive versus tracked secondary school systems · Implementation · Teacher shortage · School composition

Introduction

In this chapter, research results on the effects of system-level conditions, i.e. characteristics of the larger societal context of national education systems and policy amenable variables, are discussed. Given the state of the field of "system level effectiveness" a quantitative meta-analysis is not considered feasible and the chapter is based on a review of the relevant research literature. The selection of material does not pretend to present a representative picture and the scope of the review is limited. This chapter is to be considered in sequence to an earlier chapter, Chap. 6, in which the field of system-level effectiveness is explored conceptually. With a review of empirical results on system-level conditions, this chapter compliments the research results provided in the meta-analyses on school and teaching conditions, presented in the previous chapter.

Illustrative Research Findings on System-Level Conditions

School Autonomy

Three strands of research shed evidence on the effectiveness of increased school autonomy, as the ultimate implication of decentralization policies in education.

(a) *Studies, focusing on the effect of increased school autonomy on intermediary variables, considered as effectiveness-enhancing conditions*
Winkler and Gershberg (2000) noted that educational decentralization resulted in more qualified school principals, who exhibited a strong focus on

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_9

educational improvement. Levačić (1995) and Whitty et al. (1998) employed a similar approach, but reached different conclusions, as these authors concluded that increased school autonomy did not lead to school conditions, such as staff collaboration, which was expected to support school improvement.

(b) *Evaluation studies on school-based management and systemic reform*
Reviews by Leithwood and Menzies (1998a, b), Fullan and Watson (2000) and Bryket al. (1998), by and large showed inconclusive evidence, pointing at important interacting variables, like rigorous external accountability (Bryk et al. ibid.).
Longitudinal studies of systemic reforms in El Salvador and Nicaragua, by Jimenez and Sawada (1998) and King and Özler (1998), reported only partial effects of decentralization and increased school autonomy on student achievement.

(c) *Studies based on international comparative assessment surveys*
Another approach to study school autonomy and student achievement from a comparative point of view was employed by Walberg et al. (2000) and Wößmann (2003). Walberg and his co-authors analysed effects of decentralization policies in 14 countries. Despite their overall negative conclusion about the effect of decentralization, they did find a positive significant effect of school autonomy in hiring their staff. In his analysis of data from TIMSS 1995 and data from OECD's (1998) "Locus of decision-making", Wößmann (2003) concludes that the enhancement of school autonomy and educational decentralization, in a system with central examinations, are likely to increase student achievement. Fuchs and Wößmann (2004), using PISA data, largely corroborated these results. In their OECD study on school factors related to quality and equity, Luyten et al. (2005) report that schools with greater autonomy in the personnel management domain tend to have higher mean reading literacy scores for their students. However, as the authors note, this effect is reversed when controlling for the student composition of the school. All in all, the outcomes on the effectiveness of different aspects of school autonomy are not quite stable across studies, The evidence that enlarged autonomy in some areas of school functioning is more effective when combined with external accountability is a conceptually and practically challenging outcome, which has been partially supported in analyses on the PISA 2009 and 2012 datasets (OECD 2010a, b and 2014).

Accountability

Bishop (1997) shows, on the basis of TIMSS data, that countries with standard-based examination systems, on average, perform better than countries that do not have such examinations. Wößmann (2000), and Fuchs and Wößmann (2004), confirm the effect of examinations. Rand News Release (2000) attributes achievement gains among American states to the intensity of accountability

systems. A similar conclusion is drawn by Carnoy and others (2003). Woessmann et al. (2009) find positive effects for various kinds of evaluation procedures; the only exception is "assessment for grouping". Positive effects are reported for external exit exams, assessment for promotion, assessment for (school) comparison, the principal monitoring teachers' lessons and external inspection of teachers' lessons. They also report some beneficial effects of accountability and evaluation practices on *equity*. Teachers' subjective student assessment has a positive effect on equity, but external exit exams show a negative association. (ibid., p. 84).

School effectiveness studies have emphasized the effects of monitoring student progress, e.g. Scheerens and Bosker (1997), Willms and Somers (2000) and Scheerens et al. (2007). At the same time, other authors draw attention to potential negative side effects of high stakes testing and harsh accountability policies (e.g. Sacks 1999; Cibulka and Derlin 1995). Theoretically, the expected beneficial effects of evaluation and monitoring can be associated with system theories regarding cybernetics, research findings with respect to school performance feedback (Kluger and DeNisi 1996; Visscher and Coe 2003) and concepts of organizational learning and reflective practitioners, Argyris and Schön (1974).

School Choice

To some extent, the research evidence on demand-driven steering overlaps with results on school autonomy in general and more particularly with comparisons between private and public schools. Autonomous schools will be more likely to view their environment as a market place, and cater for the favours of the consumers of education.

Chubb and Moe (1990), among the early advocates of more school autonomy, base their arguments on the alleged benefits of private over public schools. Their study reveals that in private schools, the higher, distant authorities like school boards and supervisors have less decision-making powers than in public education. Another difference they point at is that private schools have more flexibility in personnel policies. The procedures to dismiss teachers and other school personnel are less complex and less time-consuming. Also, according to Chubb and Moe, private school principals tend to have more teaching background than public school principals. They are less interested in administrative duties than their public school counterparts, and more interested in educational matters. They conclude with a strong plea for school autonomy: "Autonomy has the strongest influence on the overall quality of school organization of any factor that we examined. Bureaucracy is unambiguously bad for school organization" (p. 183).

Chubb and Moe's work has drawn strong support, but considerable criticism as well (e.g. Glass and Matthews 1991; Smith and Meier 1995; Witte 1990). Witte (1990) was among the first to reject their findings on methodological grounds. He argued that the effects attributed to autonomy of school might in fact be caused by other school variables that were not taken into account. Moreover, he criticized the

data analysis procedures employed by Chubb and Moe, for those tended to overestimate the actual effects in the study. Smith and Meier (1995) analysed the study's theoretical claims, methods, results and conclusions, and reviewed data about the performance of school choice programs in other countries. They conclude that the available evidence does not support Chubb and Moe's case for vouchers. These conclusions are corroborated by meta-analyses on private education and voucher systems in the United States (McEwan 2000) and Latin America (Somers et al. 2004).

Besides research guided by studying the possible advantages of private education and vouchers programs in education, a large number of studies have explicitly focused on autonomous schools and their effects on student achievement over the past two decades. Most of these studies concern evaluations of reforms taking place in the United States, but also in Australia, New Zealand and—to a considerably smaller extent—in other countries as well. Research in this field often goes under the label of school-based or site-based management, or is presented under the slightly narrower term "school-based budgeting". Early review studies on the effects of these programs, presented by Malen et al. (1990) and Summers and Johnson (1996), however, provided little evidence for the notion that school-based management is effective in increasing student performance. Most of the studies were found not to be statistically rigorous, and the evidence of positive results of school-based management was either weak or non-existent (Summers and Johnson 1996).

Private Schools

The issue of school choice is related to the provision of private schooling in a country. Alleged benefits of private schools have been the subject of debate in many industrialized and developing countries. At least in industrialized countries, private schools are expected to provide better quality education than public schools. However, as private schools are funded to a large extent by fees from parents and students, students in these schools generally come from more advantaged families. At the same time, private schools are generally entitled to select students for admittance. Both the facts raise the question as to what degree private schools have an advantage in fostering high student achievement that is independent of differences in student intake. In this respect one can point to studies that indicate that differences in outcomes between public and private education become much smaller (or even disappear) if student intake characteristics are included in the analyses (McEwan 2000; Somers et al. 2004). These findings lead to the conclusion that the presumed advantages of private schooling, in terms of student outcomes, are far from clear. Moreover, critics argue that privatization leads to increased levels of segregation between students with different socio-economic backgrounds (Luyten et al. 2005).

In their analyses, based on PISA 2000 data (reading literacy performance) Luyten et al. (2005) compared independent private schools, government-dependent private schools (publicly financed, privately managed) and public schools. Their results are as follows:

"Levels of autonomy differ between private, government dependent private and public schools. Independent private schools enjoy more autonomy in all domains of decision making (personal management, learning and instruction, planning and structures and human and financial resources). According to the decision domain, government-dependent private schools fall between independent private and public schools. With regard to approving students for admittance to school, formulating the school budget and deciding on budget allocations, as well as the appointment and dismissal of teachers, government-dependent schools are very much like independent private schools. Concerning decisions on teachers' starting salaries and salary increase, the determination of the course content and student disciplinary policies, government-dependent private schools show great resemblance to public schools.

Compared to other school types, public schools are less favorably endowed with educational and physical resources, employ teachers with a (comparatively) low morale, and are characterized by poorer school climates. Independent private schools report the most favorable resources and climate, followed by government-dependent private schools. In most countries teaching and learning in public schools takes place under less advantageous conditions than in private schools. These differences in school conditions imply differential educational opportunities for students attending different school types.

However, despite these disadvantages, in half of the participating countries, there are no significant performance differences between students in public schools as compared to students in independent private schools. For the remaining countries students in independent private schools outperform students from public schools. The school composition again plays the most significant role in these performance advantages: independent private schools lose their performance advantage in all countries, once student and school characteristic are taken into account. This is confirmed by the few countries in which public schools significantly outperform independent private schools: controlling for student and school characteristics show that this is largely attributable to a more favorable intake" (ibid., p. 84).

Stratification (Tracked, Versus Comprehensive School Systems in Secondary Education)

Research studies are fairly consistent in showing negative effect of separate tracks; see Brunello and Checchi (2006), Luyten et al. (2005). These negative effects show up with respect to quality (performance levels) and equity, (SES determinacy of student achievement). Some studies find this negative association for highly

stratified educational systems only with respect to achievement levels (Luyten et al. 2005), other studies only for equity (the results from PISA 2009 and 2012 that are mentioned in the next section).

Luyten et al. (2005) summarize the results of their analyses on PISA 2000 data as follows:

"PISA 2000 also reveals that comprehensive education systems are not always more equitable in terms of variation in performance in reading literacy among students. There was no statistically significant difference between the average standard deviation in student performance of the countries with comprehensive education systems and the countries with the highest degree of institutional differentiation. However, countries with differentiated education systems are less equitable in terms of the impact of student socio-economic background on performance. Perhaps surprisingly, the PISA 2000 findings show more solid evidence for integrated, comprehensive school systems being high performers rather than champions of equity" (ibid., p. 62).

Results from PISA 2009 and 2012

PISA 2009 (Reading Literacy Performance)

Volume IV of the report on PISA 2009, (OECD 2010a, b) with reading literacy as the core performance domain is titled; "What makes a school successful; resources, policies and practices". The report provides information on the association of a number of system- and school-level variables with student achievement, more in particular results concerning *selectivity and stratification (concepts of horizontal and vertical differentiation)*, *decentralization and school autonomy, choice, accountability*, *educational resources (i.e. teacher salaries)* and *school climate*.

The concept of vertical differentiation was coined to refer to the differentiation of 15-year old students across levels (grades) in educational systems, and to grade repetition. Horizontal differentiation refers to the degree of stratification of school structures, in terms of different school categories and age of first selection. The percentage of selective schools in the system was added as a third indicator of horizontal differentiation. The results indicate that strong vertical differentiation in terms of grade repetition is negatively associated with performance. In 24 OECD countries and 27 partner countries, schools with more repeaters tended to achieve lower scores, after adjustment for student background. Grade repetition was also associated with low equity in the sense of high SES determinacy of performance.

Horizontal differentiation appeared to be unrelated to performance but tended to have a negative association with equity. School-level indices of horizontal differentiation, transfer of students to other schools because of low achievement and ability grouping appeared to be negatively associated with performance across countries. Country by country analyses showed a mixed pattern of positive and

negative associations of these variables with performance. School systems with low levels of student transfer tended to have more school autonomy in the domain of curriculum and assessment.

School autonomy, free school choice and accountability arrangements are key features of school governance, addressed in the OECD report in question. School autonomy in the domain of curriculum and assessment appeared to be positively associated with performance, while this positive association was not found with respect to autonomy in the use of school resources. Yet, in countries that had specific features of accountability in place, notably the posting of achievement of data publicly by schools, school autonomy in the domain of resources did show a positive association with performance.

Across countries, school choice did not show a relationship with performance. Within countries, schools indicating a stronger level of competition tended to de better, but this association appeared to be highly dependent on the socio-economic level of the school intake. This might be interpreted as a sign that high competition stimulates social segregation in schooling. Likewise the existence of private schools appeared to be unrelated to performance across countries, while within countries positive association depended mostly on the socio-economic status of the students.

Standard-based examinations appeared to be positively associated with performance across OECD countries. On an average, countries with standard-based examinations scored 16 points higher on reading performance than countries that do not have these. No association was found for the use of standardized tests, across countries; within country analyses pointed out that several countries did show a positive association. Use of assessment was positively associated with equity in the sense of relatively low SES determinacy of performance. Country by country analyses provided a somewhat mixed pattern of positive and negative associations as far as assessment is concerned.

Across countries, a positive association between educational resources, i.e. level of teacher salaries, and performance was found. Within school systems this association appeared to be strongly dependent on the socio-economic background of the students. Provision of pre-school education had a positive association with equity. Resource-related variables like instruction time, teacher/pupil ratio and expenditure did not show a positive association across OECD countries, but it did for all countries in PISA. The explanation that is suggested is that in industrialized countries resources do not dramatically vary between countries, and within countries, resources do not differ strongly between schools.

PISA 2012 (Mathematics)

In the PISA studies, variables that have an interpretation as system-level conditions are mostly based on country-aggregated data from responses collected at school level, from school principals. Summary results in this section are based on the PISA 2012 report, OECD (2014, vol. IV). Associations of the variables in question with

student performance can be computed at the country level and at the school level. The country-level correlations, shown in the second column of Table 9.1, although adjusted for an overall measure of wealth (GDP per capita), are to be considered as rather crude measures. When the relationship of these variables are computed for each country separately, within schools, by means of multi-level analyses, more elaborate adjustment for student background characteristics and other school characteristics are possible. The figures shown in the third column reflect the OECD mean of the adjusted school-level effects of these variables, expressed as the performance difference associated with a unit change in the independent variables. The overview in Table 9.1 indicates that generally sizeable associations found at the between country level are not reproduced when coefficients are based on within country between school measures, after adjustment for other variables. The only variable where there is a consistent country and school-level association is "teacher shortages". It should be noted that the relatively high correlations for the two indicators of horizontal differentiation (number of school types and age of first selection for secondary education) are computed with the SES impact on mathematics performance as the dependent variable. These coefficients are interpreted as measures of equity rather than quality. Horizontal differentiation had no significant association with mathematics performance. The information collected in Table 9.1 also shows that these variables computed at school level, within countries, are only significantly associated with performance in small minorities of countries, while the number of positive associations is not much higher than the number of negative associations. The main substantive results on the influence of system level policy amenable variables on student achievement (the first column of the table), show a sizeable country-level correlation for school autonomy in the curricular domain. It should be noted that this variable is based on principals' comments on the freedom that schools have in choosing text books and assessment methods. This interpretation is closer to autonomy in the domain of classroom management and instruction than to an interpretation in terms of centralized or decentralized curricula. Quality of resources shows a sizeable between country correlation with achievement and teacher shortage a sizeable negative correlation, both associations having the expected signs. The country-level correlations for accountability and teacher training are insignificant and sometimes, counter to expectations, show negative signs. The coefficient on opportunity to learn in academic mathematics, suggests that the content covered makes an important difference, within and between countries (OECD 2014, p. 150).

Direct and Indirect Effects of System-Level Variables

Direct and indirect effects of system-level variables were explored in a study by Scheerens et al. (2014). The choice of system and school-level variables was guided by the notion that the school-level variables could roughly be interpreted as school-level implementation of the system-level variables. As such the study was

Table 9.1 Relationships between variables that have a system-level interpretation, at country- and school-levels, source PISA 2012, volume IV, OECD (2014)

	System-level correlation with math performance or ESCS impact on performance*	Performance difference per unit change in independent variable, OECD average	Number of OECD countries with a significant positive association	Number of OECD countries with a significant negative association	Number of OECD countries with non-significant association
School autonomy					
Curriculum	**0.58**	0.05	1	1	32
Resources	0.00	**−4.2**	3	2	29
Choice					
Competition	−0.02	−0.01	2	1	31
Private	0.14	1.3	5	1	28
Accountability					
Data public	−0.21	**2.6**	1	1	32
Ext authority	**−0.34**	−2.1		1	33
Written feedback	0.24	−0.09	2	2	32
Horizontal differentiation					
Number of school types	**0.65***				
Age of first selection sec.	**0.63***				
Resources					
Quality educational res.	**0.58**	−0.04			34
Teacher salaries	0.31				
Teacher training					
University qualification	−0.15	0.09	4	2	28
Teacher shortage	**−0.41**	**−1.8**		3	31
Professional development	0.01	0.0	1		33
OTL formal mathematics		**48.9**	33		0 (1 missing)

Note Figures in the second column are country-level correlations. Figures in the third column are performance differences per unit change in the independent variable, obtained from multi-level analyses, controlling for student background and other school characteristics. Figures in bold indicate statistical significance at the 0.05 level. OTL is "opportunity to learn"

*The dependent variable is not math performance but the impact of ESCS (socio economic status) on math performance

described as an attempt to test assumptions of indirect causation in the hierarchical model described in Chap. 1 (Fig. 1.3) School-level data and student achievement results were obtained from the PISA 2009 database. Information on system-level variables was obtained from other sources, among others, OECD, Education at a Glance, 2010 (OECD 2010a, b). The key policy amenable variables that were chosen were: school autonomy, accountability, school choice, stratification and learning time. For each of these variables operationalizations at system and school levels were used. Starting out from the selection of system-level policy amenable variables, the choice of the school variables was determined by the notion that system and school-level variables may have a substantive meaning at both levels. School autonomy, for example can be inferred from formal regulations at the system level, but can also be measured as the way autonomy is experienced at school level. In the literature this has been indicated as "mirroring" of higher and lower level conditions, and also as "vertical consistency" between variables at different levels (Scheerens and Bosker 1997; Creemers 1994). Positive associations between system-level and school-level conditions could be seen as an implementation, negative associations, possibly as compensation at the lower level for lack of regulation from the higher level (Scheerens et al. 2014).

The effect variable in this study was reading literacy performance of 15 -year old students, as measured in PISA 2009. The index of economic social and cultural status (ESCS) was used as a covariable, defined at individual student level and at school level (school average ESCS). Data from 15 countries could be used, because of the fact that only for these the information on the system-level indicators was available.[1] The analyses to discern direct and indirect effects of country-level variables on achievement were done using structural equation modelling (SEM), a general technique for the evaluation of path models (Lee 2007).

One of the most striking outcomes was the strong influence that adjustment for SES, and particularly mean SES had on the association of the system-level variables with reading literacy achievement.

Without SES adjustments, whether having a central examination had a significant positive association with student performance (0.35). When the SES adjustments were used this positive association disappeared completely. A similar kind of difference was observed for autonomy of resources. Autonomy over school resources had a negative association of 0.24 when average SES was not included in the analysis, but almost completely vanished when school average SES was included as a coavailable. Direct effects of system-level variables that "survived" the SES adjustment test were autonomy of the curriculum (positive association) and number of school types (negative association). The high SES effects, and the substantial influence when SES is adjusted for, should be taken with caution, however. Analyses by Harker and Tymms (2004), Pokropek (2014) and Marks (2015) show that these

[1]Australia, Belgium, Check Republic, Denmark, Estonia, Hungary, Iceland, Italy, Germany, Korea, Norway, Portugal, Slovenia, Spain and the United Kingdom.

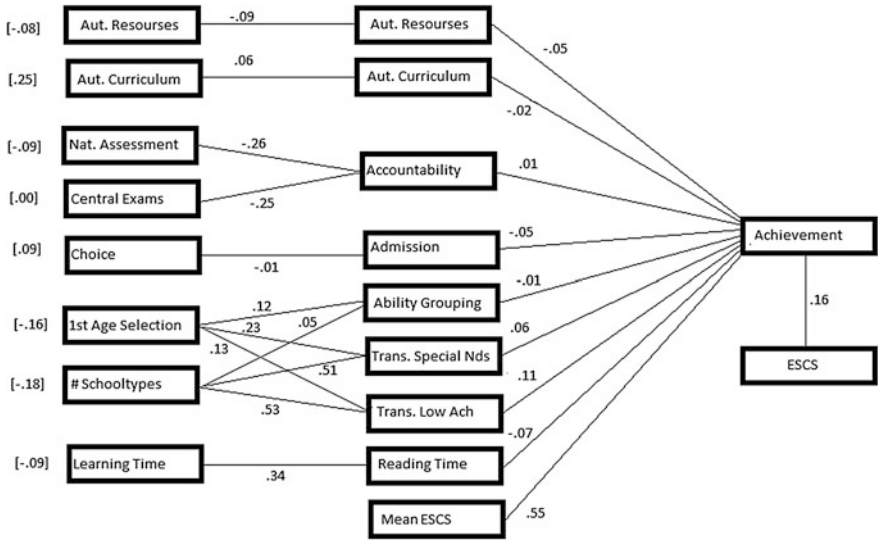

Fig. 9.1 Direct and Indirect effects of system-level variables on reading literacy achievement, adjusting for individual and school mean SES. The numbers on the *left-hand side* of the figure are the direct effects of system-level variables on achievement. From Scheerens et al. (2014)

composition effects may be statistical artifacts, due to unreliability of the SES measurements.

Figure 9.1 (from Scheerens et al. 2014) shows the complete path model; with all direct effects from the system and school-level variables, when adjustments are made for SES (indicated as ESCS[2]). Direct effects of system-level variables are shown at the utmost left of the respectively variable labels.

Some of the effects of the system-level characteristics on school-level variables are only modest, but the majority of the effects are more than 0.10 (or less than − 0.10). The following effects appeared moderately strong (the effect sizes are nearly identical in the models with and without effects of ESCS:

1. Presence of national assessments on accountability at school level (−0.26)
2. Presence of central examinations on accountability at school level (−0.25)
3. Age of first selection on ability grouping (0.12)
4. Age of first selection on reference of students to special needs schools (0.23)
5. Age of first selection on transferring students to another school because of low achievement (0.13)
6. Number of school types on reference of students to special needs schools (0.51)
7. Number of school types on transferring students to another school because of low achievement (0.53)

[2]ESCS is the Index of Educational, Cultural and Social Status.

8. Intended instruction time at system level on scheduled instruction time for reading at school level (0.34)

The associations between school-level variables and student achievement are quite small, but, except for the influence of school-level accountability, all statistically significant, given the large number of units (students). ESCS has a sizeable effect when defined at student level and a strong effect as school-level ESCS (0.55).

The sampled values of indirect effects were computed as the products of the sampled path coefficients involved. Approximately half of the indirect effects were significant, but the magnitude of these effects were small. Indirect effects associated with stratified school systems had relatively the strongest, or rather "least small" effects.

The authors (ibid., 2014) interpreted these results in terms of support for direct control (positive associations) of the system-level variables, or support for limited malleability, strong influence of ecological conditions and "loose coupling":

"Results supporting direct control and malleability
The results confirm the state of the art in the research literature in underlining the negative association of strong stratification of school systems at secondary level and student achievement. The positive interpretation of these results would support comprehensive schools and heterogeneous grouping of students. Secondly, results support the positive influence of school autonomy in the domain of learning and instruction. Only in the case of learning time a straightforward confirmation of implementation was found, in the sense of a sizeable positive association between statutory learning time and actual learning time at school.
Results supporting loose coupling, limited malleability and ecology
School average ESCS was, by far, the school level variable that showed the highest association with student achievement. Moreover, including this variable dramatically diminished the impact of one of the most promising system level levers for educational reform, namely accountability policies, measured in this study as the existence of external standard-based examinations and of national assessments. The adjusted effect of examinations that was found differs from positive findings in the literature, e.g. Bishop (1997) and Woessmann et al. (2009). Also the effect of "choice" as a system level policy changed considerably when taking ESCS at individual and school level into account. But, as it was stated before, these large means SES effects should be interpreted with caution.

The effects of the school level variables were generally small and sometimes with a negative sign that was contrary to expectations (e.g. with respect to learning time).

Due to the very small size of the associations between the school variables and student achievement, indirect effects, defined as the products of the coefficients between system and school variables and school variables and achievement, were practically negligible. This finding in itself largely falsifies implementation in the sense of indirect causation".

Dynamic Effects of System-Level Variables

Luyten and Scheerens (2014) analysed changes in system-level variables, some of which related to educational policy, others more indicative of the social context at large, in relation to changes in reading literacy performance of countries that participated in the PISA 2009 study.

Analyses were conducted on the PISA 2000 and PISA 2009 datasets. For details on the PISA study, the reader is referred to the OECD reports on PISA 2000 and PISA 2009, (OECD 2000, 2010a, b). The number of participating countries in PISA 2000 was 43, while 75 countries (of which 34 OECD member countries and 41 so-called partner countries) participated in 2009.

Next to the PISA datasets, a range of different sources were used to obtain characteristics of factors (structural and policy arrangements) defined at national system level. Lack of data availability in all countries that participated in PISA constrained the number of countries that could be used in the analyses on changes in the system-level variables and their relationship with change in reading literacy achievement.

The system-level variables (contextual and educational) are indicated in Table 9.2.

Table 9.2 Description of system-level variables, from Luyten and Scheerens (2014)

System variables
Demography
Urban population
Population density
Population aged 0–14
Population aged 15–64
Economic statistics
GDP per capita
Gini index
Child poverty
Tax wedge*
Unemployment benefits
Educational Expenditure
Teacher salaries
Expenditure as percentage of GDP
Expenditure per student
Expenditure on child care
Expenditure on pre-school
Education statistics
Instruction time
Decision-making locus
Student–teacher ratio
External examinations

*The tax wedge measures how much the government receives as a result of taxing the labor force

Correlations of Changes in System-Level Characteristics with Changes in Reading Performance Per Country (2000–2009)

Changes between 2000 and 2009 in reading the literacy achievement were computed, as well as changes on system-level characteristics for the same period (2000–2009) or a slightly different time interval. Next, the changes in system characteristics were correlated with the changes in country average reading literacy achievement. Most of these associations failed to show a statistically significant correlation with changes in reading performance. The exceptions were the *percentage of the population living in urban areas* ($r = 0.350$) and the *ratio of students to teaching staff* ($r = -0.380$). An increase in the percentage of the population living in urban areas appeared to coincide with an increase in the national reading average. The negative correlation with regard to the *ratio of students to teaching staff* indicates that the national reading average increases if the ratio decreases over time (i.e. fewer students per teacher).

Correlations of System-Level Characteristics Between Years

The correlations between years of system-level characteristics appeared to be (very) strong. This implies that countries with relatively high scores on the first measurement are likely to show high scores on the second measurement as well. The "weakest" correlation (0.627) related to expenditure on primary and secondary educations as a percentage of GDP. For intended instruction time and external examinations, the correlations were between 0.70 and 0.80. For percentage population aged 15–64, unemployment benefits, ratio of students to teaching staff the correlations were higher than 0.80 but still below 0.90. For the nine remaining system characteristics, the between-years correlation was above 0.90. A striking outcome is the limited amount of change over time for nearly all variables involved in the analyses. This is illustrated by the high between-years correlations for nearly all variables involved. This goes for both the PISA reading averages (correlation equals 0.95) and most explanatory variables.

When discussing these results the authors compare their findings with recent publications from McKinsey et al. (2010) and OECD (2010a, b), which present an optimistic view on malleable progress in education. On this they say: "Our results are generally less encouraging for a reformist educational agenda. Apart from evaluation and assessment practices (based on school level information, not discussed in this summary, JS) there was hardly any consistency to be found in the variables that characterized countries that had improved relatively strongly between 2000 and 2009. But, perhaps even more fundamentally, our results underline stability (over a period of 9 years) rather than change in both reading literacy outcomes and associated school and system characteristics. Furthermore across the board

change scores in system and school characteristics showed small effect sizes, in terms of correlations with changes in reading literacy performance. Exceptions are the *percentage of the population living in rural areas* (0.35), *student teacher ratio* (−0.38), *availability of computers* (−0.38), *ESCS-index* (−0.29), *Frequency transfer of students to another school because of low achievement* (−0.28), *Student behavior problems* (0.29) and *Use of assessments for comparing to national performance* (0.33). Of this small list of medium size correlations, no less than three show an unexpected sign (computers, ESCS and transfer). Some other variables (change scores) showed sizeable correlations as well, but did not reach statistical significance due to the relatively small number of countries: *Decisions taken at school level* (−0.36), and *Expenditure on pre-school services as % GDP* (0.28)" (Luyten and Scheerens 2014).

Discussion

System level educational effectiveness is a relatively young field of study. A handful of studies carried out in the research tradition of school improvement and systemic reform showed mixed indirect and direct effects of school autonomy on educational performance. All in all effect sizes concerning enhanced autonomy and "school based management" from this research tradition were small or totally absent. Results from recent waves of PISA (2009 and 2012) consistently show positive effects of the curriculum autonomy (OECD 2010a, b, 2014); where curriculum autonomy is defined at the level of teachers' freedom to make choices about teaching methods and instructional assessment. The way curriculum autonomy is defined is important, because these results do not seem to speak directly to the debate on content standards and (centralized) core curricula, which is currently actual in the United States (e.g. Schmidt et al. 2010), and on which expectations on improving performance are highly strung,

Large-scale studies, which looked at the effects of test-based accountability and examinations, presented mixed results. Some authors consistently report positive effects (Woessmann et al. 2009), Bishop (1997), while others were not able to replicate these positive effects (e.g. Luyten et al. 2005). Some of the results indicate that the effects of evaluation and accountability policies depend on average country wealth and average socio-economic status of the students, showing that, when these variables are accounted for, the effect totally disappears (Scheerens et al. 2014). Results on the effects of school choice hardly ever showed any positive effects. Results comparing private and public schools showed repeatedly that seemingly positive effects of the former disappeared when the socio-economic background of the students was taken into consideration. The one area in which system-level conditions were repeatedly shown to make a difference is the degree of structural differentiation of secondary school systems, comparing categorical versus comprehensive systems. It was repeatedly shown that comprehensive systems perform better, both in terms of quality (performance levels) and equity (the degree to which

educational performance is less determined by socio-economic background); e.g. (Causa and Chapuis 2009; Luyten et al. 2005, Brunello and Checchi 2006). Results from PISA 2009 and PISA 2012 confirm the advantage of less "horizontally differentiated" school systems.

With reference to the conceptual model of educational effectiveness that was described in Chap. 1 and part I of this book, it is important to note that practically all the research that was cited in the current chapter was based on direct associations between a particular system-level factor and student performance. The rather modest or absent direct effects reported have implications for expectations on being able to disentangle more complex causally mediated associations, or indirect effects. Prospects of finding significant indirect effects are bleak, when combining the evidence on the direct effects of system-level factors with the relatively small effects of school-level variables reported in other chapters. Results of the study that had included indirect effects of system-level variables corroborated the relevance of the structural differentiation of school systems, particularly as far as the number of school categories were concerned. Next, this study (Scheerens et al. 2014), showed little evidence of indirect causation and little support for implementation.

The explorations concerning system level analyses of change of countries average performance on PISA and changes in effectiveness-enhancing conditions showed fewer instances of significant change than stability. There was a remarkably strong consistency in scores on both outcomes and malleable conditions over a 9 year period. Moreover, very few significant correlations between changes in effectiveness-enhancing conditions and changes in country average achievement could be noted. These analyses, based on data from PISA 2000 and PISA 2009, where reading literacy was the central subject matter area, indicated that despite a few interesting exceptions, there was surprisingly little change in the majority of countries. This might be seen as another piece of evidence that underlines the limited malleability of educational systems, in this case not only underlining few significant associations between "causes" and "effects", but also indicating that educational reform is mostly a rather slow process.

References

Argyris, C., & Schön, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.

Bishop, J. (1997). *The effect of national standards and curriculum-based exams on achievement*. Cornell University. Center for Advanced Human Resource Studies.

Brunello, G. & Checchi, D. (2006). *Does school tracking affect equality of opportunity?* IZA Working Paper. Bonn.

Bryk, T., Thum, Y., Easton, J., & Luppescu, S. (1998). *Charting chicago school reform*. Boulder, CO: Westview Press.

Carnoy, M., Elmore, R., & Siskin, L. (Eds.). (2003). *The new accountability. High schools and High-Stakes Testing*. New York/London: Routledge Falmer.

Causa, O., & Chapuis, C. (2009). *Equity in student achievement across OECD countries: An investigation of the role of policies*. OECD Economics Department Working Papers, No. 708. OECD Publishing. doi:10.1787/223056645650.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets and American schools*. Washington, DC: Brookings Institute.

Cibulka, J. G., & Derlin, R. L. (1995). State educational performance reporting policies in the U.S.: Accountability's many faces. *International Journal of Educational Research, 23*(6), 479–492.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Fullan, M., & Watson, N. (2000). School-based-management: Reconceptualizing to improve learning outcomes. *School Effectiveness and School Improvement, 11*(4), 453–473.

Fuchs, Th., & Wößmann, L. (2004). *What accounts for international differences in student performance? A re-examination using PISA data*. CESifo Working Paper No. 1235. University of Munich, Munich, Germany.

Glass, G. V., & Matthews, D. A. (1991). Are data enough? Review of Chubb and Moe's "Politics, Markets and America's Schools". *Educational Researcher, 29*(3), 24–27.

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*, 177–200.

Jimenez, E., & Sawada, Y. (1998). *Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program.* Working papers series on impact evaluation of education reforms no. 8. Washington, DC: World Bank.

King, E. M., & Özler, B. (1998). *What's Decentralization Got To Do With Learning? The Case of Nicaragua's School Autonomy Reform*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, 13–17 April, 1998.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin, 119*(2), 254–284.

Lee, S.-Y. (2007). *Structural equation modelling: A Bayesian approach*. Chichester: Wiley.

Leithwood, K., & Menzies, T. (1998a). Forms and effects of school-based management: A review. *Educational Policy*, *12*(3), 325–246.

Leithwood, K., & Menzies, T. (1998b). A review of research concerning the implementation of site-based management. *School Effectiveness and School Improvement, 9*(3), 233–285.

Levačić, R. (1995). *Local management of schools: Analysis and practice*. Buckingham: Open University Press.

Luyten, H., & Scheerens, J. (2014). *Analysis of change in national reading literacy averages: Correlations with changes in system level and school characteristics. Manuscript*. Enschede: University of Twente.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B., & Steen, R. (2005). *School factors related to quality and equity: Results from PISA 2000*. Paris: OECD. Available from: http://www.pisa.oecd.org/dataoecd/15/20/34668095.pdf

Malen, B., Ogawa, R. T., & Kranz, J. (1990). What do we know about school-based management?: A case study of the literature—A call for research. In W. H. Clune & J. F. Witte (Eds.), *Choice and control in American education (Vol. 2): The practice of choice, decentralization and school restructuring* (pp. 289–342). London: Falmer Press.

Marks, G. N. (2015). Are school-SES effects statistical artefacts? Evidence from longitudinal population data. *Oxford Review of Education, 41*(1), 122–144.

McEwan, P. J. (2000). *Comparing the effectiveness of public and private schools: A review of evidence and interpretations* (National Center for the Study of Privatization in Education Occasional Paper No. 3). New York, NY: NCSPE, Columbia University.

McKinsey and Company-Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*.

OECD. (2000). *Skills for live. First results from PISA 2000.* Paris: OECD Publishing.

OECD. (2010a). *PISA 2009 results: What makes a school successful? Resources, policies and practices* (Vol. IV). Paris: OECD Publishing.

OECD. (2010b). *Education at a glance.* Paris: OECD Publishing.

OECD. (2014). *PISA 2012 results, volume IV, what makes schools successful? Resources, policies and practices.* Paris: OECD Publishing.

Pokropek, A. (2014). Phantom effects in multilevel compositional analysis: Problems and solutions. *Sociological Methods and Research*, 1–29.

Rand News Release. (2000). Explaining achievement gains in North Carolina and Texas.

Sacks, P. (1999). *Standardized minds*. Cambridge: Perseus Books.

Scheerens, J., & Bosker, R. J. (1997). *The Foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., Luyten, H., Glas, C. A., Jehangir, K., & Van den Bergh, M. (2014). System level indicators. Analyses based on PISA 2009 data. Internal report. Enschede: University of Twente.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organisation and Management.

Schmidt, W. H., Cogan, L. S., & McKnight, C. C. (2010). Equality of opportunity: Myth or reality in U.S. schooling. *American Educator, Winter, 2010–2011*, 12–19.

Smith, K. B., & Meier, K. J. (1995). *The case against school choice: Politics, markets, and fools*. Armonk, NY: Sharpe.

Somers, M.-A., McEwan, P. J., & Willms, J. D. (2004). How effective are private schools in Latin-America. *Comparative Education Review, 48*, 48–69.

Summers, A. A., & Johnson, A. W. (1996). The effects of school-based management plans. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's Schools: The role of incentives* (pp. 84–91). Washington, DC: National Academy Press.

Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School Effectiveness and School Improvement, 14*(3), 321–350.

Walberg, H. J., Paik, S., Komukai, A., & Freeman, K. (2000). Decentralization: An international perspective. *Educational Horizons, 78*(3), 155–165.

Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education: The school, the state and the market*. Buckingham: Open University Press.

Willms, J. D., & Somers, M.-A. (2000). *Schooling outcomes in Latin America*. Report prepared for UNESCO-OREALC and the Laboratorio Latinoamericano de la Calidad de la Educación (The Latin American Laboratory for the Quality of Education).

Winkler, D. R., & Gershberg, A. I. (2000). *Education decentralization in Latin America: The effects on the quality of schooling* (LCSHD Paper Series No. 59). Washington, DC: World Bank.

Witte, J. F. (1990). *Understanding high school achievement: After a decade of research, do we have any confident policy recommendation?* Paper Delivered at the 1990 Annual Meeting of the American Political Science Association. San Francisco.

Wößmann, L. (2000). *Schooling resources, educational institutions, and student performance: The international evidence*. Working Paper No. 983. Kiel, Germany: Kiel Institute of World Economics.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics, 65*(2), 117–170.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2009). *School accountability, autonomy and choice around the world*. Cheltenham, UK/ Northampton, MA, USA: Edward Elgar.

Chapter 10
Other Syntheses of Empirical Findings

Keywords Meta-analyses of school and teaching effectiveness · Effect sizes · School leadership effects · Time effects · Effects of evaluation as school and classroom level · Effects of assessment · Effects of feedback · Effects of opportunity to learn · Curriculum-related factors · School organizational factors · Climate factors

Introduction

The meta-analyses presented in Chap. 8 are one contribution among many to synthesize the results of empirical educational effectiveness research. The work by Hattie in this area merits special attention, as it consists of meta-analyses, not just of individual studies but of hundreds of meta-analyses (Hattie 2009). In this chapter our results, as presented in Chap. 8 will be compared to some of these other meta-analyses. In addition, reference will be made to three additional "single factor" meta-analyses that we carried out at the University of Twente, and these results will be compared with other meta-analyses, which had addressed these factors. In a final section, some results from international assessment studies will be referred to as another relevant piece of evidence. Still, it should be admitted that attempting to chart the knowledge base on educational effectiveness can never be more than providing a "glimpse" at a certain point of time, as the aim is a moving target.

Comparison with Other Multifactor Meta-Analyses

School effectiveness research is mostly field research. From the perspective of applicability, this can be seen as an advantage. Another way to express this would be to say that school effectiveness research will tend to have high-ecological validity. There is a natural connection between school improvement and applying the results of school effectiveness research, as the malleable factors identified in research are likely candidates as levers for improvement (a more extensive

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_10

discussion on the relationship between school effectiveness research and school improvement will be provided in a subsequent chapter, Chap. 11). The question "what works best in schooling" could be answered by (a) considering the set of factors on which a fair consensus among reviewers exists (see the overviews presented in Part I of this book), and (b), by rank ordering these variables according to the average effect size reported in meta-analyses. Any attempt at this kind of synthesis should be seen as tentative, because of the noted variation in effect sizes across meta-analyses, and the fact that it is not possible to capture a moving target, as new results are continuously added to the knowledge base. Nevertheless an attempt at such a tentative synthesis will be made by putting together main results from Marzano (2003), Scheerens et al. (2007) and Hattie (2009), see Table 6. Hattie's results are based on syntheses of numerous meta-analyses for each variable. In a few cases, there was not a straightforward match with variables that were included in Hattie's synthesis of meta-analyses, and somewhat specific operationalizations were chosen; these are marked and explained in the legenda of table. The variables mentioned in the overview by Marzano are taken as the starting point and rank-ordered from high to low in their association with student achievement. It appears that the original rank ordering by Marzano is preserved when one would venture to average the effect sizes across the three meta-analysis (not shown in table). The effect sizes are rendered in terms of the *d*-coefficient.

Of course the labels of the variables are quite general. In Chaps. 3 and 5, the range of specifications that is behind these general labels are made explicit. Tables included in these give more flesh and blood to the broad meaning of the variables mentioned in Table 10.1. At the same time, even the general labels provide a relatively clear idea on what aspects of school functioning should be optimized in order to enhance student performance. Opportunity to learn basically refers to a good match between what is tested or assessed in examinations, and the content that is actually taught. Instruction time may be expressed in a more global sense as officially available or allocated learning time or more specifically as "time on task",

Table 10.1 Rank ordering of school effectiveness variables according to the average effect sizes (*d*-coefficient) reported in three reviews/meta-analyses

	Marzano (2003)	Scheerens et al. (2007)	Hattie (2009)
Opportunity to learn	0.88	0.30	0.39[a]
Instruction time	0.39	0.30	0.38
Monitoring	0.30	0.12	0.64
Achievement pressure	0.27	0.28	0.43[b]
Parental involvement	0.26	0.18	0.50
School climate	0.22	0.26	0.34
School leadership	0.10	0.10	0.36
Cooperation	0.06	0.04	0.18[c]

[a]Operationalized as "enrichment programs for gifted children"
[b]Operationalized as "teacher expectations"
[c]Operationalized as "team teaching"

or "academic learning time". Monitoring may include various types of school-based evaluations, like school-based review, school performance feedback, or school aggregate measures of formative assessment at classroom level. Parental involvement might mean the actual involvement of parents with school matters, or the policies by the school to encourage parents to be involved. Achievement pressure refers not only to the school policies and practices that make use of achievement results and performance records, but also to more climate-like and attitudinal facets of fostering high expectations of student performance. School climate generally refers to good interpersonal relations at school, but often more specifically to "disciplinary climate" and the fostering of an ordered and safe learning environment. On school leadership , many specific connotations are used. Instructional leadership appears to be the most frequently used and successful interpretation in this literature. Cooperation in general terms, often measured with proxy's like the number of staff meetings, usually has a relatively weak to negligent association with student performance. Only when cooperation is explicitly task and result oriented somewhat larger effect sizes are found (cf. Lomos et al. 2011). When the rank ordering of these results is further contemplated, it appears that curriculum variables (opportunity to learn and learning time) have a relatively strong position. Monitoring could be seen as part of this curricular "syndrome", but could also be seen as a broader performance lever, which might include teacher appraisal, and schools being part of accountability schemes. The first four highest ranking factors are all to do with a focus of the primary process of teaching and learning at school. The lowest four factors are organizational measures, or "secondary processes". In the school improvement literature, variables like staff cooperation and school leadership are overrated for their importance, when one considers the quantitative evidence on performance effects. An orderly school climate is more like an organizational condition that is directly supportive of the primary process, in the sense that it is about creating a safe and productive learning atmosphere.

In the domain of teaching effectiveness, some of our results, reported in the previous chapter are compared to meta-analyses by Seidel and Shavelson (2007) and Hattie (2009). The results are summarized in Table 10.2.

Table 10.2 Results from recent meta-analyses

	Scheerens et al. (2007)	Seidel and Shavelson (2007)	Hattie (2009)
Time and OTL	0.08	0.04	0.17
Classroom management	0.10		0.26
Structured teaching	0.09	0.03	0.30
Teaching learning strategies	0.22	0.22*	0.35
Classroom climate	0.12	0.04	0.27
Feedback and monitoring	0.07	0.01	0.33

Coefficients are correlations; as Hattie presents effect sizes in terms of *d*, these were divided by two to arrive at comparable estimates.

*Teaching learning strategies was defined as "domain specific processing" in the study by Seidel and Shavelson

For comparability purposes some variation in the concepts used by the various authors were sometimes brought under one heading. For example, with respect to feedback and monitoring at classroom level from Hattie's results an average effect of 0.33 (correlation) can be computed.[1] Scheerens et al. (2007) and Seidel and Shavelson found effects of 0.07 and 0.01, respectively. Using structured teaching , as an overarching label, Hattie reports an effect size comparable to a correlation of 0.30 for direct teaching. Creemers and Kyriakides found an effect size of 0.17 for "quality of teaching", while Scheerens et al. and Seidel and Shavelson (ibid.) report effects of 0.09 and 0.02, respectively. Variables associated with quantity of teaching, time on task and opportunity to learn show effect sizes of 0.17 (Hattie, p. 184), but 0.08 and 0.03 in the studies by Scheerens et al. (2007) and Seidel and Shavelson (2007). A more extensive table with comparisons of results from meta-analyses is provided in Scheerens et al. 2011, p. 144.

According to established scientific standards, the effect sizes for the key teaching variables are medium when one considers the results by Hattie and small when one considers the other meta-analyses. One of the explanations Hattie (2009, p. 202) offers for the differences in effect sizes between his results and those by Seidel and Shavelson is that these latter authors have used only studies that controlled for student prerequisites. This could be seen as a sign that the more Europe-based studies used stricter quality controls in selecting studies, and might therefore have more credible results. The other explanation might be that effect sizes in the USA, Great Britain and Australia are higher, due to greater variability in processes and outcomes.

Results of "Single Factor" Meta-Analyses

In the period between 2010 and 2014, a series of additional meta-analyses were conducted with my colleagues at the University of Twente (Scheerens 2012, 2014; Hendriks and Scheerens 2013; Hendriks et al. 2014; Hendriks 2014). The variables addressed in these meta-analyses were time, evaluation and assessment and school leadership. The results of the computed-weighted effect sizes (Fisher's *Z* coefficients) are summarized in Table 10.3. For further details, the reader is referred to the original publications.

These effect sizes are even lower than those that were found in our earlier multivariable meta-analyses reported in the previous chapter. To give further perspective to the interpretation results are compared to results of other meta-analyses of the same variables.

Three overviews of effect sizes, concerning instruction time, school leadership and evaluation/feedback, are presented in tables.

[1]This is the average of effect sizes found for "feedback" (p. 173), "frequent testing" (p. 178) and "formative evaluation" (p. 181), expressed as a correlation.

Table 10.3 Results of recent meta-analyses Scheerens and others (2010–2014), Hendriks (2014)

Factor	Effect size	Number of effect measures
Learning time	0.05*	16
Homework at individual level	0.04*	19
Homework at class level	0.06***	12
Evaluation at school level	0.07**	7
Evaluation at class level	0.07***	15
Assessment	0.05	7
School leadership	0.06***	28

Note Effect sizes are Fisher's *Z* transformation of the correlation coefficient: *significant at 0.05, **significant at 0.01, ***significant at 0.001

The results on meta-analyses on learning time (Fraser et al. 1987; Scheerens et al. 2007; Creemers and Kyriakides 2008; Hattie 2009; Marzano 2003; Kidron and Lindsay 2014; Scheerens et al. 2014), vary (in terms of correlations), between 0.02 and 0.40 (see Table 10.4).

Table 10.4 Results of meta-analyses of time effects

Meta-analysis by	Time described as	Mean effect size (Cohen's *d*)	Mean effect size (correlation coefficient *r*)	Number of studies	Number of replications
Fraser et al. (1987) (1)	Instructional time	$d = 0.36$	$r = 0.18$		
Fraser et al. (1987) (2)	Engaged time	$d = 0.83$	$r = 0.38$	7827	22,155
Fraser et al. (1987) (2)	Time on task	$d = 0.88$	$r = 0.40$		
Scheerens et al. (2007)	Learning time	$d = 0.31$	$r = 0.15$	30	111
Creemers and Kyriakides (2008)	Quantity of teaching	$d = 0.33$	$r = 0.16$	18	
Hattie (2009)	Time on task	$d = 0.38$	$r = 0.19$	100	136
Hattie (2009)	Decreasing disruptive behaviour	$d = 0.34$	$r = 0.17$	165	416
Marzano (2003)	Classroom management	$d = 0.52$	$r = 0.25$	100	
Kidron and Lindsay (2014)	Increased learning time programs	0.03 (math) −0.04 (literacy)	$r = 0.06$ $r = -0.08$	7	
Scheerens et al. (2014)	Learning time	$d = 0.10$	$r = 0.05$	7	

Concerning leadership, average effect sizes of meta-analyses by Scheerens and Bosker (1997), Witziers et al. (2003), Marzano et al. (2005), Chin (2007). Robinson et al. (2008), Creemers and kyriakides (2008), Hattie (2009), Scheerens et al. (2007) and Scheerens (2012) are listed in Table 10.5. Effect sizes are low to very low, with one outlying result, which is the meta-analysis on transformational leadership by Chin.

Review studies and meta-analyses concerning evaluation, assessment en feedback are, among others, those by Black and Wiliam (1998), Hattie and Timperly (2007), Shute (2008), Hattie (2009) and Wiliam (2011). In Tables 10.6, 10.7 and 10.8, result from the following meta-analyses has been summarized: Fuchs and Fuchs (1985), Kim (2005), Hattie (2009), Burns and Symington (2002), Bangert et al. (1991) in Kluger and DeNisi (1996) (Tables 10.6, 10.7 and 10.8).

Again these results show sizeable differences between the various meta-analyses, for each concept. The results on evaluation and feedback, presented in Tables 10.6 and 10.8 are considerably higher than for time and leadership (except in the meta-analyses that we carried out ourselves). The overall higher effect size for evaluation and feedback is probably due to the fact that most of the results on the

Table 10.5 Summary of results from meta-analyses on school leadership; effect sizes are rendered as correlations between school leadership and student achievement

Meta-analysis by:	Leadership concept	Mean effect size (correlation)
Scheerens and Bosker (1997)	School leadership	$r = 0.04$
Witziers et al. (2003)	School leadership	$r = 0.02$
Marzano et al. (2005)	Generalized school leadership	$r = 0.25$
Chin (2007)	Transformational leadership	$r = 0.49$
Robinson et al. (2008) (1)	Instructional leadership	$r = 0.21$
Robinson et al. (2008) (2)	Transformational leadership	$r = 0.06$
Creemers and Kyriakides (2008)	School leadership	$r = 0.07$
Hattie (2009)	School leadership	$r = 0.18$
Scheerens et al. (2007)	School leadership	$r = 0.06$
Scheerens (2012)	School leadership (indirect effect models)	$r = 0.06$

Table 10.6 Results of meta-anlayses on *formative evaluation*

Variable description	Effect size (*d*)	Reference
Formative evaluation	0.70	Fuchs and Fuchs (1985)
Formative evaluation	0.39	Kim (2005)
Formative evaluation of programs	0.90	Hattie (2009)
Formative evaluation followed by an intervention	1.10	Burns and Symington (2002)
Digital pupil monitoring systems	0.06	Faber and Visscher (2014)

Table 10.7 Results of meta-analyses on assessment and testing

Variable description	Effect size (*d*)	Reference
Prestatie toetsing	0.39	Kim (2005)
Frequent toetsen	0.23	Bangert-Drowns et al. (1991)
Assessment	0.10	Hendriks et al. (2014)

Table 10.8 Results of meta-analyses on *feedback*

Variable description	Effect size (*d*)	Reference
Feedback with cues	1.10	Hattie (2009)
Feedback with reinforcement	0.94	Hattie (2009)
Feedback, computer assisted	0.52	Hattie (2009)
Feedback, overall	0.26	Bangert-Drowns et al. (1991)
Feedback, overall	0.38	Kluger and DeNisi (1996)
Feedback + correct response	0.43	Kluger and DeNisi (1996)
Feedback on changes earlier tasks	0.55	Kluger and DeNisi (1996)
Feedback and goal setting	0.51	Kluger and DeNisi (1996)

evaluation, assessment and feedback variables were computed in micro-level studies at the classroom level.

Results from International Comparative Studies

In IEA studies and PISA, school, classroom and student-level background variables from context questionnaires provide measures that can be associated with student performance. In most studies, the school and student-level context variables show a fair match with those addressed in school effectiveness research. This match is of course a deliberate strategy, as one of the purposes of the international studies is to provide policy-relevant explanations on performance differences between schools and countries, which is very similar to the "what works" mission of school effectiveness research. As an overarching re-analysis and overall review on "what works across countries", based on these international assessment studies has not been carried out, to my knowledge, some miscellaneous study results will be briefly reviewed, before some tentative general trends will be formulated.

Bosker (1997) carried out a secondary analysis of the IEA Reading Literacy Study, on which the basic report was published by Postlethwaite and Ross, in 1992. His results, when combining data from 27 countries, (100 schools per country) are summarized in the following citation (Bosker 1997; Scheerens and Bosker 1997, pp. 254–259).

"Both context indicators, *public/private* and *rural/urban,* show a positive association with adjusted school effects in reading; showing advantages for private

and urban schools. From the input indicators, *class size* has a small, and meaningless, positive effect and parental involvement has a clear positive effect (0.08). From the school process variables two achievement press variables (*focus on higher order problem solving skills* and *focus on reading*) have significant but small (0.02) positive effects. The *consensus & cooperation* indicator has a significant but small (−0.02) negative effect. The climate indicator shows a somewhat higher association (0.04). The other school process variables have estimated effects that are, statistically speaking, not discernible from zero. Of all teacher/classroom process variables only one has (an unexpected) negative effect: −0.02, namely the effect of time for reading. All in all the model for the international data does poorly, with only 9 % of unique variation between schools accounted for by the educational effectiveness variables".

In a secondary analysis of PISA 2000, focused at reading literacy performance, Luyten et al. (2005) looked at the impact of a number of malleable school variables, related to school resources, school climate and school policies. The impact of these school variables on reading literacy performance, expressed as the percentage of between school variance explained by these categories of variables was 2.1 % for the school resources, 2.2 % for the school policies and 7.7 % for the climate variables. In comparison, the percentage of between school variance explained jointly by student socio economic background at individual and school level was 51 % (resources), 50.1 % (policies) and 35.8 % (climate). These authors also presented data on the number of countries in which specific school variables were associated statistically significant with reading literacy performance. The variable that reached statistical significance in relatively most countries was the index of disciplinary climate; this variable was significant in 11 of the 39 countries. Among the school resources variable "proportion of teachers with a third level qualification in language of assessment" had the highest number of statistically significant associations at country level, but this number was limited to just 4 countries. A similar number of countries (4) scored statistically significant on the variable "Students' achievement is considered for school admission", as relatively the most frequent school policy variable reaching statistical significance at country level.

A final noteworthy result from this study was the finding of the relatively large joint effect of the climate variables and school composition, expressed as school average socio economic status of the students. The OECD average for this component was 31.1 % of the between school variation. In some countries, this joint effect was high as 66.5 (Argentina) and 63.7 (Portugal). Scandinavian countries like Iceland (0.3 %) and Finland (6.5 %) had low joint effects.

Volume IV of the report on PISA 2009, (OECD 2010) with reading literacy as the core performance domain is titled; "What makes a school successful; resources, policies and practices". Of the variables that represented elements of the learning environment those associated with school climate (disciplinary climate and teacher student relationships) were most frequently associated with performance in country by country analyses. Other variables that showed some positive association were "positive behavior among teachers" and "parent pressure to achieve". Across countries 3 % of the student performance variation was attributable to the learning

environment, when adjusting for student background; but jointly with SES the learning environment explained 9 % of the variance. Even the most successful school variables had a significantly positive association with achievement in only less than half of the participating countries. The text of the OECD report focuses on the limited number of variables that show some effect, without mentioning the fact that a much larger set of variables did not show any significant associations.

Because the previous illustrative reviews of international studies discussed in the above were about reading literacy performance, we now turn to PISA 2007, in which science performance was the effect variable. It might be argued that reading literacy is less exclusively influenced by teaching at school than subjects like mathematics and science, and therefore might show lower effect sizes of malleable conditions at school and classroom level. In the report of the PISA 2006 edition (OECD 2007), 5 school variables had small effects on student performance, after accounting for student background conditions: ability grouping (negative effect), high academic selectivity of school admittance (positive effect), schools' posting achievement data publicly (positive effect), school average time students invested in learning science (positive effect) and school activities promoting students' learning of science (positive effect), (OECD 2007, p. 267). These results are not dramatically more positive than those that were noted for reading literacy achievement. The decomposition of total performance variation in science, as reported for PISA 2006, showed that 3.4 % of the variance was explained by identified school and system level variables, net of student background conditions.

As reported in Chap. 9, in the PISA 2012 study, where mathematics was the main assessment area, opportunity to learn, particularly the exposure to formal mathematics had a significant association with mathematics performance across countries (OECD 2014, 155, 348). Moreover, exposure to formal mathematics had a positive significant relationship with mathematics performance in 62 of the 64 countries that participated in the study, with an average effect size of 50 points on the PISA scale (ibid., p. 150).

A few other secondary analyses looked at school leadership, time and opportunity to learn, respectively. Witziers et al. (2003) analysed data on school leadership in international comparative studies and found an average effect size of 0.02 (correlation). Baker et al. (2004) analysed data from three international comparative assessment studies: PISA, 2000, TIMSS, 1999 and the IEA CIVICS study (1999). Their conclusion about looking at the effect of time at between-country level is as follows: *"As a number of studies have shown, we find that there is no significant relationship at the cross-national level between the achievement test scores and the amount of instructional time"* (ibid., p. 322). When they looked at effects of time within countries they found about an equal amount of countries that had either a positive or a negative effect of time on achievement. Cogan et al. (2001) found mostly significant correlations between curriculum subtopics, (opportunity to learn) and achievement within nations.

What can we make of these results from international comparative assessment studies? The OECD is quite explicit on the limitations of the PISA datasets, as a basis for drawing conclusions about the influence of malleable system and school

conditions. The sampling design of PISA is not primarily intended to answer questions about school effects. Response on school variables is based on information from principals who are asked to generalize over the behaviour of many teachers. Student performance depends on previous years of schooling, beyond the time the students belong to the school where data collection takes place ("…*the contextual data collected by PISA is an imperfect proxy for the cumulative learning environments of students, and their effect on learning outcomes is therefore likely to be underestimated*"—OECD 2007, p. 215). Finally, the study of school resources requires precision that might not be easily captured in the surveys (ibid.).

Given these limitations, international studies have difficulty in detecting school effects, which, even in research studies, show up as relatively small. In this way, the results from international studies can be seen as a conservative test of "what works in schooling". Variables that appear to do the best in surviving this conservative test are: opportunity to learn (match between content covered and content that is tested), disciplinary climate and use of evaluation and assessment for formative application as well as accountability purposes.

Conclusions

In recent reviews, educational effectiveness researchers take for granted that the results of educational effectiveness and school improvement research provide a solid knowledge base (Reynolds et al. 2014; Muijs et al. 2014; Hopkins et al. 2014). Hattie's results, based on hundreds of meta-analyses are impressive. Instead of worrying that "nothing works", he expresses himself as being concerned that "everything works" (Hattie 2009).

When our own results of meta-analyses are compared to other meta-analyses; however, they are, without exception, rather small. Moreover, the mean effect sizes of multi- and single-factor meta-analyses show considerable variation. In the domain of school leadership, for example, mean effect sizes vary between meta-analyses from $r = 0.02$ to 0.47. To the degree that there is a systematic explanation for this variation it seems that studies carried out by European authors show smaller effect sizes than is the case for meta-analyses from authors from the USA and Australia. Whether this hypothetical interpretation is valid remains to be seen, as the bulk of primary studies used in the Europe-based meta-analyses exists of studies from the United States. Seidel and Shavelson (2007) provide evidence that suggests that study design (quasi-experiments versus surveys) might explain differences in effect sizes. They also suggest that teaching conditions most proximal to student learning will yield higher effect sizes, but this conclusion might be confounded with study design, as these micro-level processes are more likely to be studied by means of experimental designs than organizational conditions of teaching. Our results do not replicate the moderator effect of experimental versus non-experimental study design.

In Chap. 8, the educational significance of "small" effect sizes was discussed, which lead to the suggestion that effect sizes in the order of $r = 0.10$, or $d = 0.20$ could be considered as educationally relevant. When applying this standard the results of our recent "single variable" meta-analyses on time, leadership and evaluation effects remain below educational significance.

Results on the influences of school characteristics in international assessment studies often show non-existent or very small associations; with opportunity to learn as the main positive exception (Schmidt et al. 2010; OECD 2014, vol I).

Effect sizes that may be small, large percentages of unexplained variance, sizeable proportions of the variance in the student performance explained by individual or aggregate student background characteristics, little generalizability of the established set of malleable factors across countries, internationally relatively small changes in performance results over time *and* in the malleable factors that are supposed to explain them, all these considerations urge for an expedition to explore the "dark side of the moon" in educational effectiveness. When it comes to theory, this would involve looking for mechanisms that might explain not only effectiveness, but also ineffectiveness.

References

Bangert-Drowns, R. L. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 85*(2), 89–99.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–223.

Baker D. P., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional time and national achievement: Cross-national evidence. *Prospects*, *XXXIV*, 311–334.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Bosker, R. (1997). An international comparative school effectiveness study, using literacy data (Chapter 7). In J. Scheerens, & R. Bosker (Eds.), *The foundations of educational effectiveness*. Oxford: Pergamon.

Burns, M. K., & Symington, T. (2002). A meta-analysis of prereferral intervention teams: Student and systemic outcomes. *Journal of School Psychology, 40*, 437–447.

Chin, J. M. C. (2007). Meta-analysis of transformational school leadership effects on school outcomes in Taiwan and the USA. *Asia Pacific Education Review, 8*(2), 166–177.

Cogan, L. S., Schmidt, W. H., & Wiley, D. E. (2001). Who takes what math nad which track? Using TIMSS to characterize U.S. students' eighth grade mathematics learning opportunities. *Educational Evaluation and Policy Analysis*, *23*(4), 323–341.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London and New York: Routledge.

Faber, J. M., & Visscher, A. J. (2014). Digitale leerlingvolgsystemen: Een review van de effecten op leerprestaties. (Digital pupil monitoring systems: A review of effects on student achievement results). Enschede: University of Twente.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *Special Issue of the International Journal of Educational Research*, *11*(2).

Fuchs, L. S., & Fuchs, D. (1985). *A Quantitative Synthesis of Effects of Formative Evaluation on Achievement*. Washington, DC: Distributed by ERIC Clearinghouse. http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED256781

Hattie, J. (2009). *Visible learning*. Abingdon: Routledge.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.
Hendriks, M. A. (2014) *The influence on school size, leadership, evaluation, and time on student outcomes.* Enschede: University of Twente, doctoral thesis.
Hendriks, M. A., & Scheerens, J. (2013). School leadership effects revisited: A review of empirical studies guided by indirect-effect models. *School leadership & management, 33*(4), 373–394.
Hendriks, M. A., Scheerens, J., & Sleegers, P. J. C. (2014). *Effects of evaluation and assessment on student achievement: A review and meta-analysis*. Enschede: University of Twente.
Hopkins, D., Stringfield, S., Harries, A., Stoll, L., & Mackay, T. (2014). School and system improvement. *School Effectiveness and School Improvement, 25*, 257–281.
Kidron, Y., & Lindsay, J. (2014). *The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta-analytic review (REL 2014–015)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia. Retrieved from http://ies.ed.gov/ncee/edlabs.
Kim, S. E. (2005). *Effects of implementing performance assessment on student learning: Meta-analysis using HLM* (Unpublished PhD). The Pennsylvania State University, PA.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.
Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). The relationship between departments as professional communities and student achievement in secondary schools. *Teaching and Teacher Education*, *27*(4), 722–731 (10 p).
Luyten, H., Scheerens, J., Visscher, A., Maslowski, R., Witziers, B., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD.
Marzano R. J. (2003). *What works in schools. Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.
Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.
Muijs, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H., & Earl, L. (2014). Teaching effectiveness. A state of the art review. *School Effectiveness and School Improvement, 25*, 231–257.
OECD. (2007). PISA 2006. Science competencies for tomorrow's world, Vol. I: Analyses. Paris: OECD Publishing.
OECD. (2010). *PISA 2009 results: What makes a school successful? Resources, policies and practices* (Vol. IV). Paris: OECD Publishing.
OECD. (2014). *PISA 2012 results: What makes schools successful? Resources, policies and practices* (Vol. IV). Paris: OECD Publishing.
Reynolds, D., et al. (2014). Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., Teddlie, C., & Stringfield, S. (2012). Educational effectiveness research (EER): A state of the art review. *School Effectiveness and School Improvement, 25*, 197–230.
Robinson, V. M. J., Lloyd, C., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership type. *Educational Administration Quarterly, 44*(5), 635–674.
Scheerens, J. (2012). (Ed.). *School leadership effects revisited. Review and meta-analysis of empirical studies.* Dordrecht, Heidelberg, London, New York: Springer.
Scheerens, J. (2014). (Ed.). *Productive time in education. A review of the effectiveness of teaching time at school, Homework and Extended Time Outside School Hours.* Dordtrecht: Springer.
Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Scheerens, J., Luyten, H., Glas, C. A., Jehangir, K., & Van den Bergh, M. (2014). *System level indicators. Analyses based on PISA 2009 data*. Internal report. Enschede: University of Twente.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organisation and Management.

Scheerens, J., Luyten, H., & Van Ravens, J. (2011). *Perspectives on Educational Quality. Illustrative Outcomes on Primary and Secondary Education in the Netherlands*. Research Briefs, Springer.

Schmidt, W. H., Cogan, L. S., & McKnight, C. C. (2010). Equality of opportunity: Myth or reality in U.S. schooling. *American Educator, Winter, 2010–2011*, 12–19.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3–14.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Instructional leadership and student: The elusive search for an association. *Educational Administrative Quarterly, 39*(3), 398–425.

Part III
Theoretical Interpretation and Practical Application

Introduction to Part III

Theoretical interpretation of educational effectiveness research is aimed at the identification of more basic constructs and underlying mechanisms that could explain the empirical research results. Since the research syntheses only partly supported a "positive theory of educational effectiveness", explaining "ineffectiveness" was included in the search for underlying constructs and mechanisms. In Chap. 11 the state of the art of theory formation on educational effectiveness is reviewed from this perspective. Interpretations of the rational planning model are seen as useful in explaining effectiveness and ineffectiveness. Other meta-theories, loose coupling and self-organization add more to the understanding of ineffectiveness, than to suggesting alternative effectiveness enhancing approaches. In Chap. 12, before reflecting on practical application of the educational effectiveness research findings, an extensive summary of the main results of the previous chapters is given. When it comes to application, the research evidence on educational effectiveness is, first of all, analysed from the perspective of knowledge utilization. Given the nuanced reading on the solidity of the research evidence, provided in Part II, the credibility of the research findings for users is an issue that cannot be overlooked. Next, in Chap. 12, a reading is given on the scholarly field of school improvement, which basically might be considered as the executive branch of educational effectiveness research. Approaches to school improvement are analysed from the perspective of their remaining close to the research evidence. The same kind of analyses are conducted with respect to existing prescriptive frameworks for systemic reform, with a critical outlook on extensive forms of organizational development, as compared to strategies that are more directly associated with the curriculum, teaching and learning. In Chap. 13, a case study description of quality-oriented structures and programmes in Dutch education is given. The results are illustrative of certain difficulties in combining nationally orchestrated quality improvement and high autonomy of schools and intermediary levels.

Chapter 11
Theories on Educational Effectiveness and Ineffectiveness

Keywords Rationality paradigm · Loose coupling · Self-organization · Synoptic planning · Cybernetic principle · Public choice theory · Integrated educational effectiveness models · Equifinality · Transformative teleology · Composition effects · Contingency theory · Education as a complex adaptive system

Introduction

After having presented an overall conceptual model of educational effectiveness, comprising system, school and teaching effectiveness, in Chap. 1, more specific models at each of these levels were specified in subsequent chapters. Core constructs representing effectiveness enhancing conditions were described, and specified in more detail. The chapters in Part 1 of this book provide a conceptual map of what we believe "works" in education, both in terms of general strategies, but also in concrete actions and specific behaviours. It was noted that in reviews of the state of the art of educational effectiveness research a lot of consensus was displayed on the identification of the most important effectiveness enhancing conditions. Confronting this consensus with quantitative evidence on the size of effects of these variables, however, showed far less agreement, and important differences were noted among individual studies and meta-analyses. When we compared the results of the empirical studies and meta-analyses that we carried out ourselves with results from other meta-analyses, we mostly found ourselves among the authors reporting very small effect sizes, even for variables like school leadership, teaching time and evaluation and assessment, on which high expectations exist. In the final part of this book, we try to make sense of this "cognitive dissonance" between conceptual and practical expectations on the one hand and the (perhaps) disappointingly small effect sizes on the other.

This chapter is based on two articles that were published in Scheerens (2013a, b, 2015). Permission was provided by Taylor and Frances, as the publisher of the original sources. http://www.tandfonline.com/.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_11

In this chapter, the approach is to reflect on explanatory mechanisms and theory behind the empirical findings. In order to do so it is deemed important to clarify first of all what theory means in the context of educational effectiveness. We go beyond the "what works" question, by considering how and why the identified factors work. Given the frequently small effect sizes, this perspective is complemented by also considering explanatory mechanism on why certain factors do not appear to work, or have much smaller impacts than generally expected. In short not only just theories on educational effectiveness, but also theories on educational ineffectiveness are addressed.

What Does Theorizing About Educational Effectiveness Mean?

In his seminal paper about theory construction for research on teaching, Snow (1973) discusses theory construction as a gradual process, evolving from formative hypotheses about empirical regularities to axiomatic theories. The sequence is inductive, the process starts with empirically verified facts and hypotheses and generalizations being developed from these elements. Subsequent stages are, *formative hypotheses*, *elementism* (reducing the definition of variables to the most elementary units possible), *descriptive theories and taxonomies, conceptual theories and constructs* (including procedures for construct validation), *broken axiomatic theories* (this may involve eclecticism, in other words borrowing from several more established theories) and, finally, as the highest form, *axiomatic theory,* described as having a set of primitives with the help of which all its remaining concepts can be deduced and all the remaining statements can be derived as consequences (ibid., p. 83).

As far as school effectiveness is concerned, its *formative hypotheses* had both a more research based and a practical background. The scientific basis for its formative hypotheses arose as a reaction to the outcomes of the well-known Coleman report (Coleman et al. 1966), taking up the challenge that schools did matter. The practical basis was enhancing the quality of schooling, particularly for disadvantaged students. School effectiveness research and its implementation branch of school improvement has retained this dual basis, on the one hand inquiry oriented and using scientific methods, on the other hand a movement on furthering quality and equity in education. The effectiveness concept depends on establishing means-goals associations, which can be seen as formally analyzable as cause (means) and effects (attained goals) analysis. The fact that effectiveness can be placed as an important facet of educational quality (Scheerens et al. 2011) underlines the normative context of the work, and, as far as the research approach is concerned, makes for a close resemblance to evaluation research.

Elementism, the next higher up phase of theory development, according to Snow, is about the development of key concepts and instruments for the field of study. This appears to be a relatively underdeveloped area in school effectiveness research, featuring few established instruments (exceptions are perhaps the instrument development on educational leadership (Hallinger 1984), recent work within the framework of the Dynamic Model by Creemers and Kyriakides (2008) and by Van der Grift and colleagues on systematic observation of effective teaching behaviour, e.g. Van de Grift (2014).

Descriptive theories and taxonomies have been presented for integrated multi-level models of educational effectiveness since the early nineteen nineties (Stringfield and Slavin 1992; Scheerens 1992; Creemers 1994). More recently, the Dynamic Model of Educational Effectiveness by Creemers and Kyriakides (2008) has extended this work. The conceptual framework presented in Chap. 1 of this book, and elaborated in the subsequent chapters of Part 1, follows in this modelling tradition, with the ambition to give more systematic attention to system level policy amenable variables and contextual conditions and to strengthen the theoretical foundations.

Conceptual theories and constructs bring operational variables on a higher level of abstraction. When carrying out meta-analyses one has to make decisions about uniting more specific variables under a more general label, as opposed to using several more partial and specific factors. For example, transformational and instructional leadership can be united under one general leadership label, which might be indicated as integrated school leadership (Scheerens 2012). Another example is "focused instruction" as an overarching construct for constructivist and direct teaching (Louis et al. 2010). Doing so might be motivated by the finding that both of the original strategies are about equally effective (Scheerens et al. 2007; Cobern et al. 2010, Chap. 8 of this book). Using more abstract overarching constructs could be seen as one of the possible answers to deal with the fact that two distinct measures appear to be equally effective, assuming that they are just specific instances of a more general approach, and suggesting a more parsimonious underlying principle. In the case of focused teaching this might be a clear consciousness of applying a mixture of more structured teaching and independent learning methods in a teaching session.

At the next level up, Snow distinguishes *broken axiomatic theories.* In this paper only one facet of this level of theory development will be discussed, namely *eclectisism*, as this is the only facet that seems to apply to educational effectiveness research so far. In a subsequent section, results of a recent empirical review of the use of theory in school effectiveness research will be summarized (Nordebo et al. 2009; Scheerens 2013a, b).

As Snow concludes with respect to research on teaching, so can we conclude for educational effectiveness that the stage of *axiomatic theory* has not been reached.

When discussing the process of theory and model development two other terms may come up, namely those of meta-theories and paradigms. According to Snow, meta-theories are concerned with the development, investigation or description of theory itself. Examples are specific methodologies, (e.g. randomized field trials) or families of theories. The term paradigm has an even broader scope, and is used

more retrospectively in the sense that it is only applied to the analysis and evaluation of theories after their construction. In the main section of this paper, the rationality paradigm, will be used as a family of relevant theories, and it will be placed next to an orientation that unites ideas about loose coupling and self-organization.

The Way the Conceptual Model (Chap. 1) Sets the Stage for Effectiveness and In-effectiveness

The conceptual framework depicted in Fig. 1.3 of Chap. 1 showed influences across levels in a hierarchical system. Such across-level relationships can be interpreted in terms of control, facilitation and buffering from a higher level directed at the core process at the next lower level. Depicting education in this way and qualifying the overall system as hierarchical and loosely coupled has the following implications. Despite this central idea of lower level processes being contextualized and controlled by higher levels (the vertical aspect), lower levels are also seen as having considerable discretion over their core processes, in other words possess considerable autonomy. This phenomenon sets the stage for loose coupling between hierarchical levels, and implies limitations to straightforward hierarchical control. As stated in the first chapter, the degree of higher level control versus lower level autonomy is an issue of central importance at all levels. At system level it is about effective patterns of functional decentralization, which means that, perhaps dependent on the larger context, certain patterns of centralization in some functional domain (e.g. the curriculum) and decentralization in another domain (e.g. financial management) work best. At school level it is about the degree of participative decision making, or "distributed leadership", and at classroom level it refers to the balance between strongly structured to invite didactic approaches and more open teaching and learning situations that are expected to invite self-regulated learning. Structure versus independence is a red line that dominates policy and research agendas in education. A second key element in the representation in Fig. 1 is the identification of ecological conditions as a separate class of conditions influencing educational performance. This is done by giving a more explicit place to partially controllable composition effects, and their interaction with more directly malleable variables, such as the school climate. The recognition of this kind of contextual conditions emphasizes the partiality of direct control in education, and in this way underlines the loose coupling between the hierarchical levels, but at the same time focuses the attention on a qualitatively different strand of control measures, namely those of selection, admission, grouping and matching of teachers and subgroups of students as well as on cultural aspects associated with student and teacher body composition.

The basic logic of educational effectiveness research is to investigate the effects (in terms of educational outcomes) of alternative strategies, methods and approaches. Projection of this mission in the hierarchical framework described in Chap. 1

opens the way for seeing limitations and complexities to this basic logic. These limitations are evident when it comes to the implementation of higher level policies, unspecified causal chains when several acting levels are involved, the interplay between malleable factors (the effectiveness enhancing conditions) and contextual conditions, varying patterns of control and autonomy (functional decentralization), and "equifinality", the phenomenon that different approaches may have similar effects, or depend on complicated contingencies with contextual conditions.

Implementation

The relatively recent branch of educational effectiveness that considers effectiveness enhancing factors at the level of national educational systems, has given new fuel to the debate of ensuring vertical alignment in policy measures and their implementation at lower levels (OECD 2010; Mourshed et al, 2010). The respective literatures on policy implementation and planned change in education, however, offer important insights in recognizing that implementation is more often a process of "mutual adaptation", rather than straightforward "fidelity" to system level inputs. In this respect (Ball et al. 2012) coin the term "policy enactment", which they describe as "…a dynamic and non-linear aspect of the whole complex that make up the policy process, of which policy in school is just one part" (p. 6). Mutual adaptation is also described in the international school improvement literature, e.g. Elmore and Associates (1990), Miles (1998) and Mitchell and Sackney (2000). To the degree that lower levels in the hierarchical framework have more autonomy implementation of state level policies and programs becomes more problematic.

Unspecified Causal Chains

The conceptual framework indicates many cases of causal mediation, where, for example it is expected that actions by school leaders affect the teaching conditions at classroom level by way of intermediary variables, such as teacher behaviour. Educational effectiveness modelling and research does not stand out in the specification of how indirect causation is expected to work out. In the literature on school leadership effects, for example one sees an ever varying set of intermediary conditions being chosen, without any explanation on how exactly they are expected to affect student performance (Scheerens 2012). When analysing the relatively small set of empirical studies in this area, indirect effects are mostly very small (ibid.). The same, rather chaotic picture arises when a broader set of educational effectiveness studies that have used path analysis techniques to investigate indirect effect models is reviewed (Scheerens 2014a, b). A positive exception is the study by Baumert et al. (2010) on the effects of institutional arrangements for teacher training, mediated by teacher knowledge and effective teaching.

Background Conditions and Ecology

Time and time again the importance of the social economic background of students has been shown in educational effectiveness studies. It has become standard practice to control for these background conditions, in research studies and practical applications (like school inspectorates collecting "value-added" school performance outcomes). More recent studies (Opdenakker and Van Damme 2001) have also shown the importance of socio economic status at aggregate (classroom or school) level, interpreted as indicators of social competition. Results from international studies usually indicate that the composition effect is even stronger than the individual level influence of socio-economic status (Causa and Chapuis 2009). And finally, school composition may interact strongly with certain effectiveness enhancing conditions that are seen as policy amenable, particularly the school climate (Luyten et al. 2005). This "joint effect" of school climate and school average SES would mean that groups of students with certain more "favourable" background conditions are more often exposed to a more favourable school climate.

Varying Patterns of Control and Autonomy

It should be emphasized that autonomy is not at all to be seen as an instance of "failed top down control". On the contrary, autonomy in certain functional areas is often a deliberate choice in many countries that have followed policies of de-concentration and devolution of central authority during recent decades. In Chap. 1 *subsidiarity* was mentioned as a maxim expressing that "everything that can be carried out effectively at a lower level should not be carried out at a higher level". In educational systems, teaching is usually considered as belonging to the domain of the professional autonomy of teachers. Yet, among countries, important differences exist in the degree to which teacher professional autonomy is framed by central guidelines and inputs. The concept of functional decentralization indicates that systems may be centralized in one functional domain (like the curriculum, budgetary regulations, personnel policies) while being decentralized in the other. An often seen pattern is to decentralize on the way processes are given shape, while centralizing output control. This approach is at the base of accountability policies. Despite the clarity and simplicity of the logic (free process, control outcomes) practical application often leads to protests from schools and teachers who feel that their professional autonomy is constrained by accountability requirements. This in turn may evoke counter movements from within schools or educational organizations to "empower" teachers and loosen accountability regimes. The interplay of central control and regulation ambitions and autonomy at the various action levels adds considerably to the complexity of the analysis of educational systems.

Plurifinality and Feedback

The conceptual model presented in Chap. 1 is to be seen as belonging to the category of goal oriented systems. Given the fact that it is about "educational effectiveness" makes this a rather tautological statement. In general systems theory (Von Bertalanffy 1968) mentions "equifinality" and "feedback" as two important principles that model the functioning of such goal-oriented systems. Equifinality is described as "the tendency towards a characteristic final state from different initial states and in different ways, based upon dynamic interaction in an open sytem attaining a steady state". Feedback is "the homeostatic maintenance of a characteristic state or the seeking of a goal, based upon circular causal chains and mechanisms monitoring back information on deviations from the state to be maintained or the goal to be reached" (ibid., p. 46). The concept of equifinality points at the possibility that several means may have similar effects (many ways leading to Rome). The statement that the effectiveness of different ways depends on dynamic interaction in an open system is comparable to the basic principle of contingency theory, which makes the effectiveness of managerial actions dependent on the way they fit with more basic situational characteristics (so-called contingency factors, Mintzberg 1979). Equifinality, dynamic interactions and feedback are rarely considered in educational effectiveness research studies, when they are focused at the general effect of specific factors, do not use longitudinal design and do not use recursive models (which address reciprocity and feedback).

Low Elasticity of Policy Amenable Variables

Elasticity is a term used in econometrics, among others within the context of "education production functions" (e.g. Coeli et al. 1998; Callan and Santerre 1990). In most general terms, it gives an indication of the degree of change in a causal variable in order to achieve a specific change on the dependent variable. For example it may be expressed as the percentage change in the expected test score, when a conditioning variable changes by 1 %. Another example was given in Chap. 9, when the PISA 2012 results with respect to "opportunity to learn" were cited: one point increase on the opportunity to learn academic mathematics scale was associated with a change of 50 score-points in the PISA mathematics score, across countries. When the elasticity of score production to effort is low, this means that despite considerable costs and efforts effect size will still be low. The discussion of low elasticity on educational inputs in education is often illustrated with the example of class size reduction, where considerable reduction (e.g. from 24 to 16 students per class) is required to obtain mediocre effects.

Starting out from a basically managerial and technical model, but recognizing autonomy at action levels, and loose coupling between the levels, was shown to add

a lot of complexity and uncertainty with respect to straightforward malleability. In subsequent sections more substantive arguments will be provided to review the chances of policies being more or less effective.

The Way Theory Is Being Used in an Eclectic Way in School Effectiveness Research

Scheerens (2013a, b) investigated the use of theory in educational effectiveness research, in the context of a larger review of school effectiveness research (Nordebo et al. 2009). The main conclusion based on an international review of 109 school effectiveness research studies, was that only six could be seen as theory driven. This number could be, somewhat arbitrarily, raised to 11, by including those studies that were based on models that made reference to specific broader conceptual principles. These eleven studies are listed in Table 11.1.

A striking outcome was the fact that of the 11 more or less theory-driven studies 5 are based on the models by Creemers (1994) and Creemers and Kyriakides (2008). The overall conclusion from this study was that only a small minority of school effectiveness studies was more or less driven by existing theories.

Only those approaches were selected that contained ideas on specific hypotheses or at least general ideas on why certain factors addressed in empirical research would work. Articles that just contain descriptive theories of the school, or aspects of the school context, such as those by Booker et al. (2007—Bronfenbrenner's ecological and socio-cultural theory), Calaff (2008—Phelan, Davidson and Yu's Multiple World's model), Coco et al. (2004—Social constructivism and activity theory) and Fend (1998—Fend's theory of the school) are not included in the summary.

Table 11.1 Overview of studies in which more established theory was used, from Scheerens (2007)

Reference	Theory	Country
Coates (2003)	Micro-economic theory	USA
Griffith (2003)	Quinn and Rohrbaugh model	USA
Hofman et al. (1996)	Coleman's functional community theory	Netherlands
Hoy et al. (1990)	Parson's social systems' theory	USA
Kyriakides et al. (2000)	Creemers comprehensive model	Cyprus
Kyriakides and Creemers (2008)	Dynamic model of educational effectiveness	Cyprus
Kyriakides and Tsangaridou (2008)	Creemers comprehensive model	Cyprus
Reezigt et al. (1999)	Carroll model, Creemers model	Netherlands
Stringfield et al. (2008)	Schools as High Reliability Organizations	USA/UK
Tarter and Hoy (2004)	Bolman & Deal and Hoy and Miskell as theoretical bases	USA
Van der Werf (1997)	Creemers' comprehensive model	Netherlands

The theoretical approaches that were selected are: *the dynamic model of educational effectiveness*, by Creemers and Kyriakides; addressed in various publications from the list; *micro-economic* theory (Coates 2003), the *Quinn and Rohrbaugh competing values framework* (Griffith 2003) in relation to schools as High Reliability Organizations, (Stringfield et al. 2008) and Coleman and Hoffer's theory of social capital (Hofman et al. 1996) and other conceptions of the good functioning of schools such as "organizational health" (Hoy et al. 1990; Tarter and Hoy 2004).

The Dynamic Model of Educational Effectiveness

The "dynamic model" by Creemers and Kyriakides (2008) builds on the "comprehensive model" of educational effectiveness, developed by Creemers (1994). This model has much in common with other integrated, multi-level educational effectiveness models such as those developed by Scheerens (1992), Slater and Teddlie (1992), and Stringfield and Slavin (1992). Common characteristics of these models are that they combine school-level and classroom-level factors that impact on achievement. Sometimes a level of the larger context of the school is included as well. The basic rationale of these models is to take the primary process of teaching and learning as the core starting point of development. The well-known Carroll model (Carroll 1963) is mostly chosen as a guideline for modelling the primary process, emphasizing time, opportunity to learn and quality of instruction (Scheerens 1992, pp. 24, 25), School-level conditions are seen as facilitating conditions of effective teaching factors, which leads to a specific interest in cross-level interactions (cf. Bosker and Scheerens 1994). Creemers' comprehensive model defines quality, time and opportunity as basic ideas behind factors at school and classroom level. Next it goes one step further than the other similar models by defining formal principles of educational effectiveness: consistency between activities at different levels, cohesion among units (e.g. staff), constancy (stability over time) and control (internal accountability).

The dynamic model adds several ideas to the already elaborated structure of the comprehensive model:

- a specific interest in studying development over time, not only of the dependent "effect" variables but also of the independent variables, i.e. the malleable factors as classroom, school and context level;
- consideration of nonlinear relationships between the independent and dependent variables;
- next to the interest in cross-level interactions specific attention for interrelations of factors at a specific level;
- a broad outlook on effectiveness criteria (not just cognitive outcomes);
- specific measurement dimensions of effectiveness enhancing factors.

The latter characteristic makes the model quite complex. Different measurement facets are defined for all factors: frequency, stage, quality and differentiation. My interpretation of these dimensions is that frequency stands for the quantitative

intensity of a factor, stage refers to the duration of a factor being active, quality looks like construct validity (the properties of a construct as defined in the literature) and differentiation is about the adaptive implementation of a factor.

The comprehensive and the dynamic model have stimulated a number of empirical studies. The outcomes mostly speak to the tenability of certain school-and classroom-level factors. Few studies have actually investigated the theoretical properties, like the four c's of consistency, cohesion, constancy and control. In Creemers and Kyriakides (2008, Chap. 8), a study is reported in which the measurement dimensions have been tested. Results are in the direction of supporting the diversity rather than the communality of these measurement facets. This could be called good news for recognizing the complexity of educational effectiveness phenomena, but bad news for parsimony.

Education Production Functions

Education production functions describe education outputs (e.g. results on an achievement test) as a function of effort and monetary investments, taking into account innate abilities of pupils (cf. Hanushek 1979; Monk 1992). Basically, education productions functions are identical to the regression models used in educational effectiveness research. The economic background of the production function approach is most evident from the choice of independent variables, which are usually concentrated on resource inputs of schooling (teacher remuneration, class size, teacher qualification, etc.). The basic education production function represents a model, rather than a theory, and the application present in the set of studies reviewed (Coates 2003), is an attempt at overcoming under-specification of the model, in this case by employing a more refined measure of instruction time. In the wake of macro-economic studies that have sought to examine the economic gain of countries' educational performance (Hanushek and Woessmann 2009; OECD 2010) interest in production function research might be stimulated, by trying to attribute increments of economic growth to specific production elements, like early childhood education, standard based examinations, average class size, etc. (see, for example Vermeer and Van der Steeg 2011). The potential of micro-economic theory for educational effectiveness research is not so much given by production function research on its own, but rather by applications in which the behaviour or actors, like students, teachers and principals, is studied from the perspective of their utility functions (basically, how their motivation is shaped given trade-offs between task related and self-related idiosyncratic behaviour). Interesting phenomena at school level that have been studied from this perspective are standard setting (De Vos and Bosker 1998), assigning teachers to students (Monk 1992) and school composition effects (Causa and Chapuis 2009). Macro-level interpretations, concerning combinations of autonomy and control in national education policies are given in Woessmann et al. (2009). A broad overview in which educational effectiveness research is related to micro-economic theory is given in Scheerens and Van Praag (1998).

The Quinn and Rohrbaugh Competing Values Framework

Authors like Cameron and Whetten (1983) and Quinn and Rohrbaugh (1983) discuss alternative models of organizational effectiveness. Each model is based on long standing schools of thought in organizational and management science: the idea of economic rationality and rational planning, the idea of organizations as opens systems, the human relations orientation and the idea of formalized structure, the bureaucracy. Quinn and Rohrbaugh derive four distinct models from these basic orientations, the rational goal model, the open systems model, the human relations model, and the internal process model. Each is oriented towards a specific effectiveness criterion: RG towards primary production, OS towards adaptability and responsiveness towards the environment, HR towards staff job satisfaction and IP towards formal structures and procedures. Griffith (2003) uses this framework to map malleable variables that have received empirical support in school effectiveness research according to each of these four models. For example, optimizing learning time is seen as a measure that fits RG, stimulating parental involvement as belonging to OS, participative leadership is subsumed under the HR model and creating an orderly atmosphere is seen as a measure fitting the IP approach. By means of path analysis Griffith models the effects of each of the four organizational models on student achievement. The Quinn and Rohrbaugh framework has also been used, with varying success, to model direct and indirect effects of school leadership (e.g. Ten Bruggencate 2009; for an overview of studies see Scheerens 2012). By subsuming specific effectiveness enhancing variables under four broader concepts, this approach succeeds in providing a more parsimonious conceptualization of educational effectiveness. The four orientations to organizational effectiveness can easily be interpreted as different strategies to school improvement. A next step in theory development might consist of placing the preference for a specific model in a contingency framework; hypothesizing, for example that schools that are brought under a more high stakes external accountability regime would be inclined to invest in effectiveness enhancing factors associated with the rational goal model.

Two other theoretical contributions that are part of the set of studies that was analyzed can be seen, more or less, as more specific elaborations of one of the four models or organizational effectiveness. The first is the use of the theory of social capital (Coleman and Hoffer 1987) made in the article by Hofman et al. (1996). In the study in question this theory is used as a basis to investigate the structural and value consistency between school and community. The authors found some evidence that these kinds of consistencies contribute to the explanation of the superior performance of private versus public schools. The approach is in line with the open system model, and the consistency principle as emphasized by Creemers (1994).

The second theoretical approach that is represented in the set of studies and can be associated with the Quinn and Rohrbaugh framework is the work on schools as High Reliability Organizations (Reynolds et al. 2006). High reliability organizations operate in a context where failure to attain the key goals would be disastrous. Specific characteristics are: clear and finite goals, alertness to surprises and lapses, the

imperative to identify flaws, use of powerful databases, standard operating procedures, rigorous evaluation, hierarchical structure, though collective decision making. High reliability organizations are very much in line with the internal process model, and are aligned to the rational goal model as well. Bureaucratic structuring and formalization of procedures are at the heart of this approach to school improvement.

Other Conceptions of "Good" Organizational Functioning of Schools

In the articles by Hoy et al. (1990) and by Tarter and Hoy (2004) conceptions of effective organizational functioning of schools are used that are somewhat similar to the Quinn and Rohrbaugh framework, without the explicit reference to underlying schools of thought in organizational science. Basically the approach, with reference to Parsons' system theory of educational organizations (Parsons 1961) and the organizational model of Bolman and Deal (2003), defines key facets of organizational functioning, such as structure, culture, human resources, adaptation to the environment and subsequently indicates what represent good and bad organizational performance. For example, leadership should be supportive and not down-right directive, teachers engaged rather than frustrated and internal relationships should be based on trust. More specifically, Tarter and Hoy hypothesize that school structures should be enabling, the school culture should be characterized by trust, individual attitudes should be united in a sense of collective efficacy and teachers should be oriented towards the goals of the school rather than to illegitimate self-related politics. Their empirical study supports these hypotheses. An important additional notion in these conceptions of "good" schools is the idea of alignment and consistency between facets and elements: "a healthy school is one in which technical, managerial, and institutional levels are in harmony" (Hoy et al. 1990).

The Rationality Paradigm and Its Fit to the Educational Effectiveness Knowledge Base[1]

As the overview of theory use in school effectiveness research in the previous section provides a rather eclectic picture, we will turn to the application of "meta-theories" (see the earlier references to the work of Snow) to apply a more systematic approach. In this section, the rationality paradigm will be discussed as a first meta-theory. Next, a second paradigm, for which the term "transformative

[1]Part of this section is a summary of more detailed description of synoptic planning, cybernetics and public choice as presented in Chap. 1.

teleology" is used, as suggested by Stacey et al. (2000) will be explored for its relevance in explaining issues of educational effectiveness and ineffectiveness.

The very concept of educational effectiveness is based on the rational idea of optimal goal attainment. The factors "that work" can be seen as effective means to reach educational goals. This basic idea can be formalized by describing education as a contextualized production process, using the well-known context-input-process- output model, which is at the basis of the educational effectiveness models discussed throughout this book. Educational effectiveness research is particularly interested in input and process indicators that are associated with outcome indicators that represent "realized goals". Pointing at this fundamental connection of the concept of educational effectiveness with the rationality paradigm is only the first step in using it as an explanatory basis in addressing the question "why" certain factors appear to work. The second, and for our purposes more important step, is that different interpretations of the rationality paradigm indicate different explanatory mechanisms.

The basic principles of the rationality paradigm are the following: behaviour is oriented towards preferred end states (such as realized goals and personal well-being) and optimal choice is made between alternative ways to reach the goals.

Different interpretations of the rationality paradigm are: synoptic planning, contingency theory, cybernetics and public choice theory. Each of these interpretations emphasizes certain key processes, but has imperatives for organizational structuring as well.

Synoptic Planning and Bureaucratic Structuring

The ideal of "synoptic" planning is to conceptualise a broad spectrum of long term goals and possible means to attain these goals. Scientific knowledge about instrumental relationships is thought to play an important role in the selection of alternatives. Campbell's (1969) notion of "reforms as experiments" combines a rational planning approach to social (e.g. educational) innovation with the scientific approach of (quasi-) experimentation.

A modern interpretation is the concept of high reliability organizations. This concept has been successfully applied to schools (cf. Reynolds et al. 2006; Stringfield et al. 2011; Bellamy 2011).

Contingency Theory

In organizational theory contingency theory has as its central thesis that the effectiveness of organizations depends on certain more basic and contextual conditions. This is often expressed by saying that there is no universal best way to organize; success depends on a good fit between internal organizational

characteristics, and between internal arrangements and environmental conditions (cf. Kieser and Kubicek 1977; Lawrence and Lorsch 1967; Mintzberg 1979). Contingency theory can be seen as an extension of rational planning and structuring, since the ideal of optimizing remains, be it conditional on contextual conditions, which, by the way, complicates the analysis considerably (Kickert 1979). The internal alignment of organizational conditions in contingency theory is known as the "configuration hypotheses". An application in educational effectiveness and school improvement are Comprehensive School Improvement programs (Borman et al. 2003, 2004). Such programs combine an "evidence based" rational planning approach to implementation, with a coordinated set of levers for improvement, such as teaching strategies, curricular emphases, leadership and cooperation.

Cybernetics

The third interpretation of the rationality paradigm can be metaphorically labelled as "cybernetics". Cybernetics is described as the transdisciplinary approach for exploring regulatory systems. The key mechanism consists of a sequence of evaluation, feedback and corrective action; which can be thought of as a cycle.

The practice of using evaluative information on organizational functioning as a basis for corrective or improvement-oriented action can be seen as a less demanding kind of regulation than proactive synoptic planning.. In the former case planning is likely to have a more "step by step", incremental orientation, and "goals" or expectations get the function of standards for interpreting evaluative information. The discrepancy between actual achievement and expectations creates the dynamics that could eventually lead to more effectiveness.

Public Choice Theory

A central assumption in the synoptic planning and bureaucracy interpretation of the rationality paradigm is that organizations act as integrated purposeful units. Individual efforts are expected to be jointly directed at the attainment of organizational goals. In the so-called political image of organizations (Morgan 1986, Chap. 6) this assumption is rejected, emphasizing that "organizational goals may be rational for some people's interests, but not for others" (ibid., p. 195). The fact that educational organizations consist of relatively autonomous professionals, and loosely coupled subsystems is seen as a general condition stimulating political behaviour of the members of the organization.

In public choice theory (Niskanen 1971) the lack of effective control from democratically elected bodies over public sector organizations marks these organizations as being particularly prone to inefficient behaviour, essentially caused by the leeway that is given to managers and officers to pursue their own goals, besides

Table 11.2 Interpretations of the rationality paradigm

Interpretation of the rationality paradigm	Mechanism
Synoptic planning	Proactive structuring
Contingency theory	Fit
Cybernetics	Evaluation and feedback
Public choice theory	Market mechanisms; school competition, choice

serving their organization's primary mission.[2] Creating competition and market mechanisms are seen as the remedy to overcome these problems. The alleged superiority of private over public schools is the most supportive piece of empirical effectiveness research for the claims of public choice theory, although the significance of the results in question is much debated (Scheerens 1992; Goldhaber 2000; Gorard et al. 2001; OECD 2005). A similar observation could be made with respect to decentralization and enhanced school autonomy as a lever of educational performance (Scheerens and Maslowski 2008; Luyten et al. 2005).

An overview of the four different interpretations of the rationality paradigm and their key mechanisms is given in Table 11.2.

As the overview has shown, these four interpretations of the rationality paradigm and their corresponding key mechanisms can be matched with important levers of educational performance, such as proactive planning approaches, highly formalized organizational structures like schools as high reliability organizations, Comprehensive School Reform projects, all kinds of applications of evaluation and assessment, including accountability policies and data use, and measures to make schools more autonomous and private and to stimulate free school choice. As shown on earlier occasions (Scheerens 1997) matching can also be carried out in a more detailed way, by relating key variables reflecting effectiveness enhancing conditions to each of the rationality perspectives (see Table 11.3 at the end of the next section).

What Rational Models Have to Say About Educational Ineffectiveness

Empirical studies suggest that the antipodes of effectiveness enhancing factors are associated with school failure. Stringfield (1998) mentions "lack of academic focus", "academic periods starting late and ending early", "bureaucratic leadership" and "lack of teacher assessment", as characteristics of failing schools.

These examples show negative effects of "too little" of the factors that are associated with rationality interpretations. At the same time, there are extreme

[2]A more extensive treatment of the implications of public choice theory for school effectiveness research is given in, Scheerens (1992, Chap. 2).

Table 11.3 Variables addressed in educational effectiveness research matched to four alternative interpretations of the rationality paradigm, as well as associated sources of ineffectiveness

Rationality interpretation	Global intervention strategies	Variables addressed in research	Side effects and Exaggerations, associated with ineffectiveness
Synoptic planning and bureaucratic structuring	Proactive planning Evidence-based policies Schools as High Reliability Organizations	**System level** – National curriculum planning – Infrastructure for program evaluation in education – Centralized structures, limited school autonomy – School inspection **School level** – instructional leadership – Consistency of practice – Academic emphasis and achievement orientation – Clear and fair discipline **Classroom level** – Time and opportunity to learn – Structured lessons – Monitoring and feedback	Standardized operating procedures in teaching Goal displacement "Red tape" Lack of flexibility and implementation problems
Contingency theory	Comprehensive School Reforms Differential effectiveness	**System level** – Vertically aligned educational systems **School level** – Transformational leadership **Classroom level** – Adaptive instruction – Differentiation	
Cybernetics	Accountability Organizational learning Education based meritocracy	**System level** – A broad range of accountability provisions – National assessments – School inspection – New public management: free processes control outcomes	Negative side effects of high stakes testing Resistance to assessment and evaluation Factors preventing organizational learning Evaluation apprehension

(continued)

Table 11.3 (continued)

Rationality interpretation	Global intervention strategies	Variables addressed in research	Side effects and Exaggerations, associated with ineffectiveness
		– (Institutionalization of) school self-evaluation – (Institutionalization of) school development planning – Facilities for continuous professional development of teachers **School level** – School self evaluation – School performance feedback **Classroom level** – Pupil monitoring systems	
Public choice theory	Free school choice Privatization School autonomy Competition	**System level** – Free school choice – Financial and managerial school autonomy – Privatization – Vouchers – "High stakes" accountability arrangements **School level** – Merit pay of teachers – Educational entrepreneurship **Classroom level** – Stimulating extrinsic motivation – Efficient class size – Matching teachers and students – Optimizing class size	Off-task behavior. Political processes. "Make work", Exaggerated managerial overhead

implementations of rationality models, which might be globally indicated as "too much" of them.

In the concept of schools as high reliability organizations, there is a thin line between an acceptable degree of standardization and rigorously monitored "standardized operating procedures SOP's", as the favoured focus of quality management systems (ISO and others). These latter practices cause a lot of "red tape" and might lead to goal displacement; teaching to the SOP would seem to be worse than teaching to the test!

In the domain of evaluation and feedback a whole literature exists about the negative side effects of high stakes testing and intensive external school inspection, like tunnel vision, all kind of strategic behaviour and administrative burden (cf. Ehren 2008). In less contested areas as school self-evaluation and the image of schools as learning organizations, resistance, immunization against potential criticism and barriers to organizational learning have been noted as well. As part of their theory of "organizational learning" Argyris and Schön (1978) recognize limits to organizational learning. These limits reside in behavioural patterns of the members of the organization. They describe these as "shared strategies in individual theories in use". As examples they mention considerations like:

- "Let buried failures lie.
- Keep your views of sensitive issues private; enforce the taboo against their public discussion.
- Do not surface and test differences in views of organizational problems.
- Avoid seeing the whole picture; allow maps of the problem to remain scattered, vague, ambiguous" (p. 39/40).

They also mention that such strategies may reflect deeper and more fundamental norms, strategies and assumptions:

- "Protect yourself unilaterally—by avoiding both direct interpersonal confrontation and public discussion of sensitive issues which might expose you to blame.
- Protect others unilaterally—by avoiding the testing of assumptions where that testing might evoke negative feelings, and by keeping others from exposure to blame.
- Control the situation and the task—by making up your own mind about the problem[3] and acting on your own view, by keeping your view private, and by avoiding the public inquiry which might refute your view" (p. 40).

Public choice theory and its foundation, micro-economic theory has off-task behavior and selfish motives of organizational members, as one of its foundational pillars. One could even say that the theory's basis is negative, in the sense of preventing dysfunctional organizational functioning next to positively oriented to

[3]Argyris and Schön apply these strategies to a situation where an organization is confronted by a significant developmental or strategic problem.

rational management. Market mechanisms, including competition and effective assessment strategies are invented to drive such dysfunctional behaviours out. There is correspondence with the phenomena mentioned by Argyris and Schön (ibid.) under the heading of limits to organizational learning, as micro-economic theory takes into consideration protective, defensive and egoistic reactions by members of the organization. A summary overview of the specification of the various interpretations of the rationality model, including references to ineffectiveness, is presented in Table 11.3.

Alternative Theories for Explaining Effectiveness and Ineffectiveness

The interpretations from the rationality paradigm discussed in the above go a long way in explaining the findings from educational effectiveness research. It is conceivable to propose a theory driven research agenda, in which the effectiveness of pure forms of the models, and more partial derivatives, could be tested, and where the alternatives could be pitted against one another. In this, contingency theory is a case apart, since it could be used as a meta-theory to investigate in which situations one of the three other models would work best. Implementation failures, exaggerated interpretations and undesired side effects would appear to be plausible explanations for ineffectiveness.

For several reasons it is still considered useful to go beyond the analysis of positive and negative instances of the rationality paradigm and look at alternative theories. These are the following:

1. The applications of the rationality paradigm will tend to be focused at the productivity of the operational core of the organization; and organizational structural conditions, including links to the environment, that support this; this would mean that other organizational functions, such as responsiveness to new developments in the environment, reflection on aims and goals, and cultural facets of the organization get less emphasis.
2. Rationality applications would seem to have a stronger preoccupation with instrumental and technological levers for improvement than with improving cultural conditions and motivational aspects of work; human relations and recruitment policies.
3. Rationality applications are closer to proactive and retroactive planning and structuring modes of organizational functioning than to implementation.
4. Rationality applications would tend to focus on the formal organization and have less attention for the informal organization.

Two alternative perspectives will be discussed that might provide a different outlook on educational effectiveness and ineffectiveness: schools as loosely coupled organizations, and theories about self-organization. The question whether loose

coupling provides an alternative perspective to interpretations of the rationality paradigm is open for debate. It is also associated with systems theory, incrementalism and "bounded rationality" (Lindblom 1959; Simon 1964), and as such still part of the rational paradigm. At the same time bounded rationality offers quite different orientations towards educational effectiveness and ineffectiveness, as well as certain prescriptive interpretations regarding organizational change. Self-organization is used as the central term to refer to complex interactions in organizations that emphasize "emergence" rather than control.

Schools as Loosely Coupled Systems

Weick describes "loose coupling" as an "image that coupled elements are responsive, but that each event also preserves its own identity and some evidence of its physical and logical separateness" (Weick 1976, p. 1), He goes on to say that "Loose coupling also carries connotations of impermanence, dissolvability and tacitness all of which are potentially crucial properties of the 'glue' that holds organizations together" (ibid., p. 1). In educational organizations, two of the most fundamental couplings, the one among elements of the core technology, and the other between the authority structure on the one hand and the functioning of the technical core on the other, are not particularly strong.

More specific combinations of elements in educational organizations between which loose rather than tight coupling is likely to occur are:

- intentions and actions
- yesterday and tomorrow
- top and bottom
- line and staff
- administrators and professionals
- several means leading to the same end (equifinality)
- teachers and materials
- voters and the school-board
- parent and teacher
- teacher and pupil

Schools as loosely coupled organizations would seem to be at odds with educational effectiveness models that depend on "rational planning" type of mechanisms. The concept of equifinality, (Von Bertalanffy 1968) alone seems to present fundamental problems for the causal implications of the very concept of educational effectiveness. Loose coupling could be read as an explanation for the relatively low and inconsistent effects for core factors like coordination and consensus, educational leadership and evaluation and feedback. Yet, according to Weick, loose coupling has certain advantages. Tight and loose coupling are to be seen as forming a continuum, where weaker couplings, like shared conceptual anticipations and retrospections, may nevertheless create a certain robustness and resilience of the

organization, because they contain mutations, localized adaptations and fewer costs of coordination (ibid., p. 14).

To the extent that the educational effectiveness research orientation and knowledge base emphasizes the rationality paradigm, loose coupling offers an explanation for its mediocre success: why some factors do not work optimally, variability in effect sizes across studies and meta-analyses, instability of the research findings, and basic doubts about the generalizability of the research outcomes. The analysis of implementation problems with evaluation and feedback procedures, from the perspective of loose coupling, for example provides a conceptual basis for underutilization, misuse and undesired side effects, in applications like school self-evaluation (Scheerens 2004) and accountability (Ehren 2008). To the extent that schools confirm to the properties of loosely coupled systems, low effects of leadership should not come as a surprise. Facets of loose coupling, such as none clarity of intentions, individual members of the organization pursuing different goals, and the notion that several means may lead to the same end, put question marks behind some of the basic assumptions of the educational effectiveness approach.

In his well-known article from 1976 "Educational organizations as loosely coupled systems", Weick presented an outline of a research program, based on further analytic and empirical study of "loose coupling". Elements of this program are: developing conceptual tools capable of preserving loosely coupled systems, explicate which elements are available in educational organizations for coupling, development of contextual methodology, the collection of thorough, concrete descriptions of the coupling practices in actual educational organizations, specification of the core technology in educational organizations, probing empirically the functions and dysfunctions associated with loose coupling, discover how inhabitants make sense out of loosely coupled worlds (Weick 1976, p. 18). It would be an interesting question for research on science, why this theory driven research program, does not seem to have made it, despite the pervasiveness and actual relevance. So far Weick's theory on loose coupling does not seem to have had very much influence on empirical educational effectiveness research.

Self-organization[4]

The "complexity sciences", chaos theory, dissipative structure theory and the theory on Complex Adaptive Systems (Stacey et al. 2000), look at dynamic interactions between the micro-elements of a system, trying to model these, or to come to grasp with regularities or rules that emerge. Study of complexity developed in the natural sciences; compare the work of Maturana and Varela (1980) and Pirogine (1997). Famous examples are the way molecules behave in combustion, the butterfly effect

[4]This section builds on the introduction of self-organization presented in Chap. 1.

and the discovery of patterns that are "stable and instable" at the same time, so called fractals.

Stacey et al. (ibid.) describe the philosophical background of "normal science", as compared to various interpretations of complexity theory. According to them, self-organization, when it is more than unfolding already initially given patterns, is able to create novelty and is to be seen as a different kind of causality. The mechanistic image of organizations, as implied in scientific management and the "formative teleology" of general system theory is compared to the more organic idea of "transformative teleology". According to Kaufman (1993) formative causality is self-referential: "the dynamics cause themselves as the system evolves of its own accord to the edge of chaos" (Kauffman cited by Stacey et al. p. 117).

Given our purpose to connect educational effectiveness research and its knowledge base to more established theory, particularly management theory in the broadest sense, it should be noted that, according to some interpretations of complexity theory, we have moved beyond the scientific paradigm and what Stacey et al. indicate as rationalist teleology. In fact the efficiency perspective is considered as a far too limited criterion to indicate what organizations are aiming for. Likewise "normal" causal analyses, or analyses focused at optimizing means-end relationships, are not considered in these interpretations. Instead, as cited above, self-organization is seen as a different kind of causality, and the ultimate intension of organizational development is sometimes indicated in terms of survival, but also in terms of "expressing identity" (Goodwin 1994, cited by Stacey et al. 2000, p. 119). If educational effectiveness research and school improvement would embrace the ideas of the "transformative teleology" interpretation of complexity, this would indeed mean a paradigm shift (compare: Harris et al. 2013; Scheerens 2013b).

Before looking at social science and educational applications, some further characteristics of the way the dynamic interactions between the micro elements of a system are seen in applications of Complex Adaptive Systems theory will be briefly reviewed.

Importance of Initial Conditions

The initial characteristics of the micro elements that interact are considered of great importance, according to Stacey et al. (2000) the diversity in the elements is a pre-condition for interactions that may lead to novelty and creativity.

Patterns of Stability and Instability

The dynamics lead through phases of stability and instability, progressing from one "state space" of equilibrium (also indicated as "attractor") to another, culminating in

what is indicated in chaos theory as a "strange attractor". Strange attractors are described in various ways, like "patterns which are repeated, but never exactly the same", and examples are given in the realm of weather prediction and heart rhythms, but no examples for social science or management applications seem to exist.

Nonlinear Development

Interaction and change are seen as evolving in a nonlinear way, sometimes allowing for "jumps" and qualitative changes in the system.

Preoccupation with Disorder

Creativity is associated with initial states of disorder and diversity. "Not only is the system restless, but its own restlessness allows it no rest. Restlessness about restlessness may increase restlessness", writes Luhman (1995, p. 50). In terms of change processes, one could say that a permanent state of "unfreezing" is considered good for innovation.

Rules Amidst Chaos

Despite of the previous point, some regularities are expected to emerge from chaos, complexity is not an aim in itself. Examples of regularities and rules that may emerge are provided in the image of "fitness landscapes", clusters of intensive "good" interactions as fitness peaks, and low intensity cluster as valleys (below, a research example will be provided in which these metaphoric concepts are made more concrete).

Non-Managed Dynamics

The interactive processes are thought of as autonomous, circular and self-referential. There is no room for an objective observer or for a controller; in social systems is only a participatory role imaginable.

Applications in the social sciences are sometimes purely metaphoric (e.g. Morgan 1986), but appear also as researchable conjectures. Stacey et al. (2000) concentrate on the interrelationships between members of an organization, while the sociologist Luhman (1995) also considers dynamics of structural arrangements.

With respect to organizations the applications from complexity theory have a focus on what is often indicated as the informal organization, of relatively "hidden" properties such as the deep level of organizational culture (cf. Schein 1985a, b), and the hidden curriculum. Stacey et al. describe the importance of this level of functioning as follows: "…systems in organizations can only function if the members weave their day-to-day interactions with each other through and around the rules of the systems they have designed" (ibid., p. 59). Though these authors expect positive developments, particularly novelty and creativity from the free interaction between the members of the organization, they are also open to the possibility that dysfunctional results and undesired consequences may occur. Next, as far as application to organizations is concerned, the focus is on describing phenomena in the realm of implementation and enactment, as any kind of management and planning is banned from the research agenda. Phenomena like "ownership" and co-construction fit the ideas on "formative teleology" quite well. In both the theories by Luhman and Kaufmann, there is attention for the issue of changing the boundaries of the organization with respect to the environment, and networking between subgroups within and outside the organization.

Before reflecting on the implications for theorizing about educational effectiveness two research applications are briefly sketched. One looks at communication patterns among school governors and principals in the context of implementation of the US No Child Left Behind project, (Daily et al. 2011). The other addresses the issue of composition effects (Scheerens 2004, 2008).

Daily et al. (2011) used complexity theory as a basis to study "how rational assumptions undergirding current reform policies limit our understanding of how policy is enacted through complex social interactions". They studied the implementation of No Child Left Behind policy measures, targeted at improving consistently underperforming schools and applied longitudinal social network modelling to illustrate how school districts could be conceptualized as complex adaptive systems. The interaction patterns were analyzed in terms of "emergence", "fitness peaks" and feedback.

The authors conclude that the study suggests that policy implementation is a complex endeavour and does not necessarily follow linear, predictable patterns as might be suggested by conventional policy assumptions (ibid., p. 26). As these patterns are considered to be decisive for the uptake of the reform, and may vary across schools, large-scale standardized approaches to educational issues may be inadequate in addressing local problems. Emergence of reciprocity and feedback, on the other hand, is associated with socio-cultural learning, and development of "ownership" of the reform. Spontaneously developing centralization (disproportional influence of district leaders) led to disconnection between district and school leadership level and was dysfunctional to the expected growth of interactions between the two levels.

A second application of ideas from complexity science is provided in the interpretation of composition effects by Scheerens (2004, 2008). Composition effects can be seen as illustrating the importance of starting conditions at the micro-level, for organizational conditions at a higher-up level. In school

effectiveness research student background conditions at individual and aggregate level are often used as control variables. Important technical progress has been made on the measurement and modelling of student background composition effects (De Van Damme et al. 2000). The way composition effects operate has facets of a non-managed process, although deliberate selection and grouping policies may be brought into play to control them. For those who like the, in my view, mystifying charm of the phrase that the whole is more than the sum of its parts, this might be applied to composition effects. A more sober statement could be that composition leads to phenomena at a higher-up level with distinct "higher-up level" interpretations, not immediately obvious from the micro-level characteristics. Scheerens (ibid.) suggests that a phenomenon like the school culture might be conceptualized as a composition effect based on the aggregation of personality traits of teachers; associating for example a culture of "high expectations" with a high proportion of teachers with and externally oriented locus of control. A final way in which composition effects reflect some of the issues that is often discussed in the realm of complexity theory is the nonlinearity and qualitative "jumps" in development and change. Some analysts, for example have addressed the question of the proportion of special needs students and thresholds for heterogeneity being still manageable by regular teaching provisions, within the framework of inclusive education, and similar critical thresholds concerning the proportion of minority students.

In making up the balance, insights from complexity theory emphasize a number of "positive" phenomena, factors in education that could be seen as enhancing effectiveness and improvement. These could be summarized as: stimulating autonomy and decentralization, exploiting and optimizing composition effects, providing space for spontaneous interaction and grouping, and in the case of externally induced reform, a keen eye for an enactment (or mutual adaptation) perspective on implementation and "ownership".

On the other hand the interactions among micro level elements of the system may also go wrong, and lead to power games, the pursuit of egoistic behavior, inertia, and resistance to sensible reform measures. Throughout this chapter attention has been given to policy failures, dysfunctional implementations and, in short, in-effectiveness. Complexity theory offers another idiom to understand particularly the dynamic process dimension of negative development, but does not provide substantive ideas and explanations about why things go wrong. To the extent that complexity theory is the antagonist of rational planning, and an engineering interpretation of education change, it is often positioned as a source of explanation of why such policies do not work, or have only limited success. Such critique could be taken more or less radically. A radical solution would be to see emergence and "self-organizing" as a message to abandon all kind of planning, management and control. A more modified view would be to think of organizations as having both controlled and self-organizing facets, with attention for both formal and informal aspects of organizing. To me this seems the preferable and more pragmatic solution. In the application of the complexity sciences to social systems, there seems to be a blind spot as far as the reality of the formal organization is

concerned. Perhaps a parallel could be made with analyses of formal and informal authority and power in organizations (Pfeffer 1978.). Even when concentrating on informal power, in this literature it is recognized that those with formal authority enter the arena with lead in their gloves.

Discussion

In this presentation two extremes of the graded continuum of theory development by Snow have obtained most of the attention, on the one hand the "basic" empirically supported knowledge base, consisting of average effect sizes of a fairly consistent set of effectiveness enhancing conditions, and on the other hand rather general or meta-theories. A middle part of the continuum, consisting of middle range theories and verified intermediary effect models has been treated in a more cursory way (see the section on the eclectic use of theory in school effectiveness research). In case a program of theory-oriented research synthesis would get off the ground, it could be an interesting job to accumulate and categorize such middle level conceptual work in a more systematic way. The usefulness of having charted the domain of meta-theories up front might help in making such endeavour more focused.

Scientific educational effectiveness research would imply that "refutations" should be seen as equally important as the confirmation of "conjectures", to paraphrase Popper. The idealistic urge to reform and improve education could get in the way of a neutral and balanced attention for the impact of malleable factors. In a review of studies about the effects of programs to prevent school drop-out, Kane (2004) concluded that "unpublished research and higher quality methods produce significantly smaller treatment differences". Favouring studies that show high impact on educational performance is a not unlikely type of publication bias. The attempt at reviewing the knowledge base on educational effectiveness in this book, presented in earlier chapters, points at large differences in the average effect sizes found across meta-analyses and small effects and little generalizability across countries found in international studies. It might be that the malleable conditions most frequently addressed in educational effectiveness research are on the small side, although still educationally relevant. This underlines the interest to know more about the limits of the rational techniques that dominate the educational effectiveness and the education reform agenda. The theories that were described in this paper provide indications about causes of ineffectiveness, in the sense of implementation failures and side effects, and collision between the realities of the formal and informal organization. This information might be used in various ways in educational practice and policy: by actively countering implementation failures and side effects, by fostering more realistic expectations on effects and effect sizes among practitioners and policy makers, and by considering alternatives levers for improvement. Weick's theory of loose coupling shows that foregoing active interference and "letting go" or resorting to socialization mechanisms could

occasionally be more effective. His theory underlines the importance of the initial training of teachers and the functionality of professional autonomy. The "Finnish approach" (Sahlberg 2011), could be seen as an alternative lever for educational reform and improvement, very much centred on teacher initial training, esteem, and professional motivation. Complexity theory shows an interest in the informal organization, and elements of unpredictability in the interactions of organization members and the emergence of new patterns of behavior, which could be functional or dysfunctional. Similarly self-organization could be used as an interpretation of composition effects, and nonlinear developments caused by drastic changes in the composition of student and staff intake.

References

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Ball, S. J., Mac Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. New York: Routledge.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 133–180.

Bellamy, RH. (2011). High reliability leadership for educational change. In *McREL: High reliability organizations in education*. Denver: McNell. http://www.mcrel.org/pdf/LeadershipOrganizationDevelopment/0121MM-HRD-Noteworthy-sml.pdf

Bolman, L. G., & Deal, T. E. (2003). *Reframing organizations: Artistry, choice and leadership*. San Francisco: Jossey Bass.

Booker, K. C., Invernizzi, M. A., & McCormick, M. (2007). "Kiss your brain": A closer look at flourishing literacy gains in impoverished elementary schools. *Reading Research and Instruction, 46*, 315–339.

Borman, K. M., Carter, K., Aladjem, D. K., &LeFloch, K. C. (2004). Challenges for the future of comprehensive school reform. In C. T. Cross (Ed.), *Putting the pieces together. Lessons from comprehensive school reform research* (pp. 109–150). Washington: National Clearinghouse for Comprehensive School Reform.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research., 73*, 125–230.

Bosker, R. J., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. Conceptual and methodological advances in educational effectiveness research. Special issue of the *International Journal of Educational Research*, *21*, 159–180 (Oxford: Pergamon Press).

Calaff, K. P. (2008). Supportive schooling. Practices that support culturally and linguistically diverse students' preparation for college. *NASSP Bulletin, 92*(2), 95–110.

Callan, S., & Santerre, R. (1990). The production characteristics of local public education: A multiple product and input analysis. *Southern Economic Journal*, 68–480.

Cameron, K. S., & Whetten, D. A. (Eds.). (1983). *Organizational effectiveness. A comparison of multiple models*. New York: Academic Press.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*(4), 409–429.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 722–733.

Causa, O., & Chapuis, C. (2009). *Equity in Student Achievement Across OECD Countries: An investigation of the role of policies*. OECD Economics Department Working Papers, No. 708. OECD Publishing. Doi: 101787/223056645650.

Coates, D. (2003). Education production functions using instructional time as an input. *Education Economics, 11*(3), 273–292.

Cobern, W., Schuster, D., & Adams, B. (2010). Experimental comparison of inquiry and direct instruction in science. In *Presentation at SREE Spring Conference*.

Coco, A., Frid, S., Galbraith, P., Gholam, M., Goos, M., Horne, M., et al. (2004). *Home, school and community partnerships to support children's numeracy*. University of Queensland.

Coelli, T., Prasada Rao, D. S., & Battese, G. E. (1998). *An introduction to efficiency and productivity analysis*. Boston: Kluwer Academic Publishers.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of education opportunity*. Washington, DC: Government Printing Office.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools. The impact of communities*. New York: Basic Books Inc.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Creemers, B., & Kyriakides, L. (2008). *The Dynamics of Educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. New York: Routledge.

Daily, A. J., Moolenaar, N. M., & Carrier, N. (2011). *Reform at the edge of chaos: Connecting complexity, social networks, and policy implementation* (manuscript).

De Van Damme, J., Fraine, B., Opdenakker, M. C., Van Landegem, G., & Onghena, P. (2000). *Effects of schools and classes upon mathematics and language achievement: The importance of group composition*. Edinburgh: Presentation at ERE.

De Vos, H., & Bosker, R. J. (1998). Micro-economic explanations for educational effects. In J. Scheerens & B. M. S. van Praag (Eds.), *Micro-economic theory and educational effectiveness* (pp. 73–104). Enschede: University of Twente.

Ehren, M. C. M. (2008). Het spel der neveneffecten; strategisch gedrag van toezichthouders en scholen. In F. L. Leeuw, J. S. Kerseboom, & R. Elte (Eds.), *Turven, tellen, toetsen; over toezicht, inspectie, handhaving en evaluatie en hun maatschappelijke betekenis in Nederland*. Den Haag: Boom Juridische Uitgevers.

Elmore, R. F., & Associates. (1990). *Restructuring school: The next generation of educational reform*. San Francisco: Jossey Bass.

Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. (Quality in education; educational research on system level conditions, school profiles and student achievement)Weinheim: Juventa.

Goldhaber, D. (2000). School choice: Do we know enough? *Educational Researcher, 29*(8), 21–22.

Goodwin, B. (1994). *How the Leopard changed its spots*. London: Weidenfeld & Nicolson.

Gorard, S., Fitz, J., & Taylor, C. (2001). School choice impacts: What do we know? *Educational Researcher, 30*(7), 18–23.

Griffith, J. (2003). Schools as organizational models: Implications for examining school effectiveness. *The Elementary School Journal, 104*(1), 29–47.

Hallinger, P. (1984). *Principal instructional management rating scale*. New York, NY: Leading Development Associates.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14*, 351–388.

Hanushek E. A., & Woessmann, L. (2009). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation* (NBER Working Paper, No. 14633). Cambridge MA, National Bureau of Economic Research.

Harris, A., Chapman, C., Muijs, D., & Reynolds, D. (2013). Getting lost in translation? An analysis of the international engagement of practitioners and policy makers with the educational effectiveness research base. *School leadership & management, 33*(1), 3–19.

Hofman, R. H., Hofman, W. H., Guldemond, H., & Dijkstra, A. B. (1996). Variation in effectiveness between private and public schools: The impact of school and family networks. *Educational Research and Evaluation, 2*(4), 366–394.

Hoy, W. K., Tarter, C. J., & Bliss, J. R. (1990). Organizational climate, school health, and effectiveness: A comparative analysis. *Educational Administration Quarterly, 26*(30), 260–279.

Kane, T. J. (2004*). The impact of after-school programs: Interpreting the results of four recent evaluations*. Working Paper. New York: WT Grant Foundation.

Kauffman, S. A. (1993). *Origins of order: Self organization and selection in evolution*. Oxford: Oxford University Press.

Kickert, W. J. M. (1979). *Organization of decision-making: A systems-theoretical approach*. Amsterdam: North-Holland Publishing Company.

Kieser, A., & Kubicek, H. (1977). *Organiation.* De Gruyter Lehrbuch.

Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers' comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4), 501–529.

Kyriakides, L., & Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183–205.

Kyriakides, L., & Tsangaridou, N. (2008). Towards the development of generic and differentiated models of educational effectiveness: A study on school and teacher effectiveness in physical education. *British Educational Research Journal, 34*(6), 807–838.

Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and environment*. Cambridge: Harvard University Press.

Lindblom, C. E. (1959). The science of muddling through. *Public Administration Review*, *19*.

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement, 21*(3), 315–336.

Luhmann, N. (1995). *Social systems*. Stanford, CA: Stanford University Press.

Luyten, J. W., Scheerens, J., Visscher, A. J., Maslowski, R., Witziers, B., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000*. Paris: OECD.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition. The realization of the living*. Dordrecht/Boston/London: D. Reidel Publishing Company.

Miles, M. (1998). Finding keys to school change: A 40-year Odyssey. In A. Hargreaves, A. Liebermann, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change.* Dordrecht, Boston, London: Kluwer.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Mitchell, C., & Sackney, L. (2000). *Profound improvement: Building capacity for a learning community*. Lisse: Swets & Zeitlinger.

Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis, 14*(4), 307–332.

Morgan, G. (1986). *Images of organizations*. Beverly Hills: Sage.

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. McKinsey and Company.

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Chicago: Aldine-Atherton.

Nordenbo, S. E., Holm, A., Elstad, E., Scheerens, J., Soegaard Larsen, M., Uljens, M., et al. (2009). *Research mapping of input, process and learning in primary and lower secondary schools.* In: The evidence base. Copenhagen: Danish Clearing House for Educational Research, DPU, Aarhus University.

OECD. (2005). *Education at a glance. OECD indicators 2005*. Paris: OECD.

OECD. (2010). *The high cost of low educational performance. An estimation of the long-run economic impact of improvements in PISA outcomes*. Paris: OECD.

Opdenakker, M., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*(4), 407–432.

Parsons, T. (1961). *Theories of society: Foundations of modern sociological theory*. New York: Free Press.

Pfeffer, J. (1978). *Organizational design*. Arlington Heights: A.H.M. Publishing Company.
Prirogine, I. (1997). *The end of certainty: Time, chaos and the new laws of nature*. New York: The Free Press.
Quinn, R. E., & Rohrbauch, (1983). Spatial models of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science, 29*, 363–377.
Reezigt, G. J., Guldemond, H., & Creemers, B. P. M. (1999). Empirical validity for a comprehensive model on educational effectiveness. *School Effectiveness and School Improvement, 10*(2), 24.
Reynolds, D., Stringfield, S., & Schaffer, G. (2006). The high reliability schools project: Some preliminary results and analyses. In A. Harris & J. H. Chrispeels (Eds.), *Improving schools and educational systems* (pp. 56–76). Abingdon: Routledge.
Sahlberg, P. (2011). *Finnish lessons*. New York: Teachers College Press.
Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.
Scheerens, J. (1997). Conceptual models and theory-embedded principles on effective schooling. *School Effectiveness and School Improvement, 8*(3), 269–310.
Scheerens, J., & Praag, B. M. S. van (Eds.). (1998). *Micro-economic theory and educational effectiveness*. Enschede/Amsterdam: University of Twente, Department Educational Organisation and Management/ University of Amsterdam, Foundation for Economic Research.
Scheerens, J. (2004). The evaluation culture. *Studies in Educational Evaluation, 30*, 105–124.
Scheerens, J. (2007). *Conceptual framework for the PISA 2009 Context questionnaires and thematic reports*. Oslo: Paper for the PISA Governing Board.
Scheerens, J. (2008). L'istruzione intesa come sistema adattativo complesso: implicazioni per gli studi sull'efficacia educativa. In F. Abbona, G. Del Re & G. Monaco (eds.), *Complessità dinamica dei processi educativi* (pp. 136–171). Milano: FrancoAngeli.
Scheerens, J. (2012). (Ed.). *School leadership effects revisited. Review and meta-analysis of empirical studies.* Dordrecht, Heidelberg, New York, London: Springer.
Scheerens, J. (2013a). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement., 24*(1), 1–38.
Scheerens, J. (2013b). The ripples and waves of educational effectiveness research: Some comments to 'Getting lost in translation. *School leadership & management, 33*(1), 20–25.
Scheerens, J. (2014a). *Productive time in education. A review of the effectiveness of teaching time at school, homework and extended time outside school hours*. Dordrecht: Springer.
Scheerens, J. (2014b). *What operational path models tell us about educational effectiveness.* Key-note address at the conference 'The international school inspection project. Modalities and mechanisms for effective schooling', 4 June 2014, Gothenburg.
Scheerens, J. (2015). Theories on educational effectiveness and ineffectiveness. *School Effectiveness and School Improvement., 26*(1), 10–31.
Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organisation and Management.
Scheerens, J., Luyten, H., & Van Ravens, J. (2011). *Perspectives on educational quality. Illustrative outcomes on primary and secondary education in the Netherlands* (148 p.). Research Briefs, Springer.
Scheerens, J., & Maslowski, R. (2008). Autonomie des établissements scolaires: des moyens à la recherche d'un objectif? *Revue francaise de pedagogie, 164*, 27–36.
Schein, E. H. (1985a). *Organizational culture and leadership: A dynamic view*. San Francisco, CA: Jossey-Bass.
Schein, E. H. (1985b). *Organizational culture and leadership: A dynamic view*. San Francisco, CA: Jossey-Bass.
Simon, H. A. (1964). *Administrative behavior*. New York: Mac Millan.
Slater, R. O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement, 3*(4), 247–257.
Snow, R. E. (1973). Theory construction for research on teaching. In R. M. W. Travers (Ed.), *Handbook of research on teaching* (pp. 77–112). Chicago: Rand McNally.

Stacey, R. D., Griffin, D., & Shaw, P. (2000). *Complexity and management. Fad or radical challenge to systems thinking?*. New York: Routledge.

Stringfield, S. (1998). An anatomy of ineffectiveness. In L. Stoll & K. Myers (Eds.), *No quick fixes: Perspectives on schools in difficulties* (pp. 209–221). London: Falmer.

Stringfield, S., Reynolds, D., & Schaffer, E. C. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the high reliability schools project. *School Effectiveness and School Improvement, 19*(4), 409–428.

Stringfield, S. C., Reynolds, D., & Schaffer, E. C. (2011). Toward high reliability public schooling. In R. H. Bellamy (Ed.), *McREL: High reliability organizations in education*. Denver: McNell.

Stringfield, S. C., & Slavin, R. E. (1992). A hierarchical longitudinal model for elementary school effects. In B. P. M. Creemers & G. J. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35–39). Groningen: ICO.

Tarter, C. J., & Hoy, W. K. (2004). A systems approach to quality in elementary schools. A theoretical and empirical analysis. *Journal of Educational Administration, 42*(5), 539–554.

Ten Bruggencate, G. C. (2009). *Maken schoolleiders het verschil? [Do School Leaders Make a Difference?]*. Enschede: University of Twente.

Van de Grift, W. J. C. M. (2014). Measuring teaching quality in several European countries. *School Effectiveness and School Improvement*, *25*(3), 295–311 (16 p).

Van der Werf, G. (1997). Differences in school and instruction characteristics between high-, average-, and low-effective schools. *School Effectiveness and School Improvement, 8*(4), 430–448.

Vermeer. N., & Van der Steeg, M. (2011). *Onderwijsprestaties Nederland* (Educational Performance of the Netherlands). Central Planning Bureau (CPB); Policy Brief, 2011; The Hague.

Von Bertalanffy, L. (1968). *General systems theory*. New York: George Braziller.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*, 1–19.

Chapter 12
Recapitulation and Application to School Improvement

Keywords School effectiveness · School improvement · Systemic reform · The educational effectiveness knowledge base · Scenarios for systemic educational reform · Strategies for evidence-based school improvement · Exposure to educational content · Evaluation · Monitoring and feedback provisions · Managing the teaching and learning programme · School leadership as meta-control · Alternative levers for school improvement · The curriculum and climate emphasis · The teacher recruitment and professional development strategy · Optimizing matching and grouping at school level

Introduction

In this chapter the balance is made up about the educational effectiveness research results that were brought together in this book and their theoretical interpretation. Next, the use and application of the "knowledge base" is addressed in discussions about research use and usability and strategies for school improvement and systemic reform. This leads to a comparison between the fields of educational effectiveness research and the study of school improvement, and to the identification of reform strategies at school and system level. What we have is a positive theory on what works in education, with a fair consensus about key levers for improvement at system, school and classroom level. But then, there are a lot of additional considerations that call for prudence in joining the "heroic journey" towards organizational excellence that some school improvement experts are envisaging. Despite broad consensus on the factors that are positively associated with student performance, there is much divergence in reviews and meta-analyses about the actual effect sizes. Despite the rationality and common sense of the conceptual models, there was a story to be told about the negative side effects and interpretations and theories that explain ineffectiveness. And finally, implementation of the research results into practice appears to be a complex endeavour, where improvement guidelines frequently go far beyond strong factual support.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_12

What Works at What Level, and "How Much"

The core part of Part 2 is the report on meta-analyses that we carried out on educational effectiveness studies conducted between 1997 and 2005. The results are divided into a part on school-level factors and a part that deals with variables at the teaching level. The choice of the variables in these meta-analyses was based on the literature reviews reported in Part 2. On the whole, effect sizes were low. At school level they were comparable to the meta-analyses reported in Scheerens and Bosker (1997). At teaching level the effect sizes were only slightly higher than at school level, but much lower than in the meta-analyses reported by Hattie (2009) and Marzano (2003).

In the school-level analyses the relatively highest effect sizes (of the order of $r = 0.15$) were found for the variables *opportunity to learn* and *effective learning time*. The combination of these two variables can be interpreted as degree of exposure to goal-relevant educational content. Curricular choices, test preparation and "instructional alignment" can be seen as the major levers for improvement.

Effect sizes of almost similar size were found for the climate variables *orderly climate* and *achievement orientation* (0.12 and 0.14, respectively). In terms of malleability these variables should be looked upon with more caution than the previous effects of exposure, however. It has been argued repeatedly that variables like achievement orientation and high expectations are ambiguous as far as causal direction is concerned. As argued by Brophy (1984), high expectations may depend on adequate information about the capacity of students to achieve at a certain level, making the expectations the effect rather than the cause. Orderly climate is a variable that is very sensitive to school and classroom composition effects. When the average socio-economic status of students is at, say middle-class level, it is easier to create an orderly task-oriented atmosphere (cf. Thrupp 2008; Luyten et al. 2005). The magnitude of the joint effect of school climate variables and school composition, as reported in the recent waves of PISA (OECD 2010a, 2014) is another piece of evidence arguing for caution in seeing these climate variables as malleable in a simple and straightforward way.

Monitoring had a relatively small effect size in this meta-analysis (0.07); in the earlier meta-analysis (Scheerens and Bosker 1997) the effect size for monitoring was 0.14. The conclusion about the average effect in the current analysis should be modified, however, since effect sizes in important subject matter areas as language and mathematics are of the order of 0.18 and 0.17, as will be argued further on in this chapter, monitoring, evaluation and feedback have a convincing theoretical interpretation as levers of educational improvement, but, unfortunately, may be plagued by implementation problems and undesired side effects as well.

School organizational variables like *leadership* and consensus and cooperation had very small effect sizes (0.05 and 0.02). The tiny effect for leadership matches the results in the earlier meta-analyses by Scheerens and Bosker (1997) and Witziers et al. (2003). It should be noted that these outcomes stand in sharp contrast to claims of the importance of leadership in the influential study by Day et al. (2009), which is one of the key supports for the MC Kinsey report on leadership (Barber et al. 2011).

Relatively small effect sizes were found for *parental involvement* (0.09) and *homework* (0.07). *Differentiation* had an effect size, as low as $r = 0.02$.

When in Chap. 10 these results were compared to outcomes in meta-analyses by Marzano (2003) and Hattie (2009) it was noted that the effect sizes that we found are generally much lower. They are more in line with results from Seidel and Shavelson (2007) and Creemers and Kyriakides (2012), which might suggest that in meta-analyses that contain a sizeable number of European studies, effect sizes are considerably lower than is the case for analyses based on studies from the United States, New Zealand or Australia. As pointed out in Chap. 10, Table 12.1, despite the differences in effect magnitudes the rank ordering of factors in terms of effect size is roughly similar across these meta-analyses, namely in the following way:

1. opportunity to learn and effective learning time,
2. monitoring and achievement pressure,
3. parental involvement,
4. orderly climate,
5. school leadership,
6. cooperation.

To this might be added that effect sizes for differentiation and individualized instruction are relatively low, in the meta-analyses by Hattie, as in ours. He reports an effect size of 0.10 (in terms of a correlation coefficient) of individualized instruction.

The results of our multi-factor meta-analysis on effectiveness enhancing teaching factors were reported at three different levels of specificity: 46 teaching variables, 15 more general factors, and finally in terms of 6 even more synthetic factors. In the latter most synthetic version the following factors were distinguished: *curriculum, classroom management, structured teaching, constructivist-oriented teaching, climate* and *evaluation and monitoring*. Of these 6 broad categories *constructivist-oriented teaching* had a mean effect size of 0.13, while *structured teaching strategy* showed a mean effect size of 0.07. *Curriculum* had a mean effect size of 0.11, *climate* of 0.09, *classroom management* of 0.08 and monitoring of 0.07. Apart from the fact that these effect sizes (expressed as correlations) were all relatively small, the most striking outcome was the fact that *constructivist-oriented teaching* had a higher effect size than *structured teaching*. When considering the 15 more general factors, the highest effect sizes were *learning to learn learning strategies* (0.21), *teacher characteristics*, such as *high expectations and constructivist beliefs about teaching* (0.15), a *cognitively challenging teaching approach* (0.13), a learning *environment that is orderly and characterized by low achievement pressure* (0.13); a *clear and structured teaching approach* (0.13) and "*active*" *teaching*, characterized by a variety of didactic approaches (0.12). The coefficient for *support* was 0.11.

At the level of the more specific teaching variables the following variables had the highest mean effect sizes (mean effect sizes are rendered as correlations, the figures between brackets are mean effect sizes for comparable variables, reported in Hattie's (2009) synthesis of meta-analyses):

Goal-directed/clear teaching 0.22 (0.37)
Cooperative learning 0.20 (0.27)
Peer tutoring 0.22 (0.27)
Meta-cognitive training 0.24 (0.35)
Scientific inquiry training 0.20
Reading and writing strategies 0.21 (0.18)
Cognitive activation 0.18 (0.32)

These results can be interpreted by comparing them to results from other meta-analyses and to theoretical perspectives about effective instruction.

Comparing to other meta-analyses shows the same overall picture as noted for the school-level variables. Our effect sizes are generally smaller than those reported by meta-analyses that are more exclusively based on studies from the USA, Australia or New Zealand (Marzano 2003; Hattie 2009), and more comparable to results of partially Europe-based studies (Creemers and Kyriakides 2008; Seidel and Shavelson 2007).

Constructivist-oriented teaching strategies, most notably related to learning to learn domain-specific skills and cognitive activation, came out relatively stronger in our meta-analyses, than did variables associated with direct teaching. Hattie and Marzano reported effect sizes of the order of 0.30 for direct teaching and mastery learning where we only found mean effect sizes of about 0.10–0.05. The only variable in the category of structured teaching approaches that did better was goal-directed teaching and learning (0.22).

When putting our results into a broader perspective it seems to be wiser to conclude that both structured, proactive goal-directed teaching and strategies that capitalize on learning to learn, meta-cognitive strategies and cognitive activation are effective. The common denominator of both strategies seems to be a readiness to reflect on didactic strategy in a structured way, while being both proactive and retroactive in monitoring educational content and processes. In this context Louis et al. (2010) coined the term "focused teaching" for an approach that combines direct teaching and constructivist elements.

When we carried out additional meta-analyses on educational leadership, time and evaluation and assessment, the result showed even smaller average effect sizes. These are summarized in Table 12.1, where they are compared with the range of average effect sizes reported in other meta-analyses of these variables.

Table 12.1 Results from recent meta-analyses (Scheerens 2012a, b, 2014; Hendriks 2014)

Factor	Effect size	Range of effect sizes in other meta-analyses
Learning time	0.05*	−0.04 to 0.40
Homework	0.05*	0.07 to 0.29
Evaluation at school level	0.07**	
Evaluation at class level	0.07***	0.06 to 0.70
Assessment	0.05	0.10 to 0.39
School leadership	0.06***	0.04 to 0.49

Note Effect sizes are Fisher's *Z* transformation of the correlation coefficient: *significant at 0.05, **significant at 0.01, ***significant at 0.001

Hendriks (2014, 221) discusses various possible reasons why these results are lower than those that were found in other meta- analyses. First of all, she mentions that the majority of studies in the three meta-analyses referred to in the above was correlational and did not have an experimental or a quasi-experimental design. As a matter of fact, in the meta-analyses on leadership and time, all studies had a correlational design. In the meta-analysis on evaluation and assessment only 4 of the 21 studies had used an experimental or quasi-experimental design. Secondly, she mentions evidence that to the degree that the primary studies included more independent variables, next to the variable of interest, effect sizes tended to be lower. And since most studies did in fact include several independent variables, this could be another explanation for the low mean effect sizes in the meta-analyses. To this could be added that it has sometimes been observed that educational research studies that use more sophisticated methodology tend to have slightly smaller effect sizes Kohn (2004), and this is also what we found in the meta-analyses on school level variables, reported in Chap. 8). As the more recent studies are better than earlier ones in this respect, e.g. by a widespread application of multi-level analysis techniques, this might be an additional explanation. Yet, these are just conjectures and not convincing reasons to distance ourselves from the results we found. In other words, from the material we analysed and for the variables we looked at it seems that the effect of conditions that are very credible as effectiveness enhancing variables are really quite small.

With respect to the influence of system-level variables it was concluded that a handful of studies carried out in the research tradition of school improvement and systemic reform showed mixed indirect and direct effects of school autonomy on educational performance. All in all effect-sizes concerning enhanced autonomy and "school-based management" from this research tradition were small or totally absent. Results from recent waves of PISA (2009 and 2012) consistently show positive effects of curriculum autonomy (OECD 2010a, 2014), where curriculum autonomy is defined at the level of teachers' freedom to make choices about teaching methods and instructional assessment. The way curriculum autonomy is defined is important, because these results do not seem to speak directly to the debate on content standards and (centralized) core curricula, (e.g. Schmidt et al. 2010), and on which expectations on improving performance are highly strung. Results from PISA (OECD 2014, vol. I) that do speak for proactive alignment of teaching content and achievement are the results on opportunity to learn (see the relatively high average within country correlation reported in Chap. 9) PISA 2012 further showed that the quality of educational resources and teacher shortages had sizeable effects (0.58 and −0.48 respectively) at the between-country level.

Large-scale studies, which looked at the effects of test-based accountability and examinations, presented mixed results, with some authors consistently reporting positive effects (Woessmann et al. 2009), Bishop (1998), and others not replicating these positive effects (e.g. Luyten et al. 2005). Some of the results indicate that the effects of evaluations and accountability policies depend on average country wealth and average socio-economic status of the students, showing that, when these variables are accounted for the effect totally disappears (Scheerens et al. 2014). Results on the effects of school choice hardly ever showed any positive effects. Results

comparing private and public schools showed repeatedly that seemingly positive effects of the former disappeared when the socio-economic background of the students was taken into consideration. The one area in which system-level conditions were consistently shown to make a difference is the degree of structural differentiation of secondary school systems, comparing categorical versus comprehensive systems. It was repeatedly shown that comprehensive systems perform better, both in terms of quality (performance levels) and equity (the degree to which educational performance is less determined by socio-economic background; e.g. (Causa and Chapuis 2009; Luyten et al. 2005; Brunello and Checchi 2006). Results from PISA 2009 and PISA 2012 confirm the advantage of less "horizontally differentiated" school systems.

In the attempt that was made in Part 2 of this book, to assess the knowledge base on educational effectiveness, a very nuanced picture was painted of "what works". In many areas there was a lot of variation in the mean effect sizes that were reported for the same variables. In our own meta-analyses and reviews of international assessment data we found generally smaller effects than were reported in earlier meta-analyses, and consequently we do not rule out the possibility that malleability of educational effectiveness is (even) smaller than is usually thought. Compare the rather optimistic tone in the recently published state-of-the-art reviews (Reynolds et al. 2014; Hopkins et al. 2014; Muijs et al. 2014). Next it should be noted that the picture is different for system-level, school-level and classroom-level variables that are considered as stimulating effectiveness. Drawing all the evidence together the following summary is the result.

At *system level* the only malleable condition that could be called established (repeatedly confirmed as a significant influence on student performance) is the *horizontal differentiation of secondary school systems* (tracked versus comprehensive school systems), with comprehensive systems doing better in terms of achievement levels and equity of outcomes. Mixed research results were noted for *school autonomy, accountability, human resources* (*teacher training effects*), *curriculum characteristics and financial and material resources.* Results are mixed in different ways. Results are sometimes positive but sometimes not, as in the case of *autonomy*, appear to be confounded with background conditions, such as average socio-economic status, or general wealth of a country, which is the case with respect to *accountability,* only occur in certain segments of the distribution, as is the case with *financial and material resources,* which make less of a difference in industrialized wealthier countries than in less developed countries. As far as teacher training effects is concerned the situation is characterized by a majority of studies in the education production function tradition showing very small effects but some promising results from recent international studies (TEDS-M[1] in particular) and a handful of national research studies. Little support is found for parent's free *choice* of schools and *private* schools (where gross effects generally disappear when adjusted for socio-economic background).

[1]Teachers Education and Development Study in Mathmeatics. Blömeke et al. (2014).

At school level, a certain consistency was observed between meta-analyses in supporting *opportunity to learn* as the condition with the relatively highest mean effect size. An *orderly managed school, with implications for discipline and effective teaching time* is the next best conglomerate of conditions, closely followed by frequent *monitoring* and *achievement orientation*. *Educational leadership, consensus and cooperation* and *differentiation* had very small average effect sizes in the studies that were reviewed.

At the level of teaching effectiveness, the research domain is much more varied, diversified and specialized than is the case for the higher aggregation levels of the national system and the school. Many results show about the same level of effect sizes; in some meta-analyses average to high, in others quite low (recall Hatties' remark that "everything works"). When variables and factors are aggregated to rather broad areas, differences are flattened out, but it is still striking that major schools of thought, like the more behavioristic structured teaching approaches and cognitivist-oriented constructivist approaches both appear to have quite similar positive effects. Perhaps in this area it is best to conclude that effective instruction is characterized by a dedicated and focused orientation on teaching strategy (containing elements of cognitivist and direct teaching approaches), a well-managed and supportive school climate, opportunity to learn and frequent use of formative evaluation and feedback. Here the empirical results that were summarized are well in line with more conceptual summaries of factors, as made by Brophy (2001) and Klieme (2012), and where core latent teaching factors are described in terms of "support and classroom management", "supportive classroom climate" and "cognitive activation" (Scheerens 2013a, b, 28).

Theoretical Interpretation of the Empirical Findings

The overview of the empirical knowledge base showed much variation in mean effect sizes from meta-analyses, very small effect sizes in our own contributions, a very mixed picture of successful and less successful system-level factors, quite modest effects of school-level malleable variables, many teaching-level variables with rather similar effects sizes and relatively strong effects of background conditions and composition effects.

The main purpose of connecting these empirical findings to theory is to see to what extent these results can be understood as incidents of more general constructs or underlying mechanisms. Given the state of the art of theory formation, as explained in Chap. 11, in the case of educational effectiveness this means taking a second look at the structure of the conceptual models that are used, more specifically the models introduced in Part 1 of this book, and see to what extent findings can be understood form the middle range and meta-theories that were discussed in Chap. 11.

Characteristics of the Conceptual Model and the Research Evidence

When starting out from the model described in Chap. 1, all of the substantively interesting associations, those considering contextual conditions and background variables, across level causation and investigation of indirect effects, were seen to be covered in empirical educational effectiveness research. Some of these associations were practically addressed as standard practice, with others being addressed relatively rarely.

Controlling for background conditions has been standard practice, and considering composition effects is happening more frequently since one and a half decade. As far as studying indirect effect models and causal mediation across levels is concerned it was noted that this happens in relatively few studies. In rare cases these indirect effect studies are driven by at least some kind of substantive rationale on how intermediate causation is supposed to work. Examples of studies that stand out positively in this respect are Heck and Moriyama (2010), Baumert et al. (2010) and Kyriakides and Creemers (2012). In most cases, however, such studies are not model-driven and just exploratory. Even in the reports one does not find much reflection on what the occasionally found significant path coefficients actually mean. The same applies to the interpretation of direct effects, particularly when factors at high level, remote from student learning are associated with student performance. Micro-economic theory has a straightforward interpretation of the effect of a country having a standard-based national examination on student achievement, namely that students' motivation to do well is strongly enforced. But when it comes to school-based accountability, where school performance is assessed, the interpretation of an expected positive relationship with student performance is less simple. Carnoy et al. (2003) found evidence for external accountability requirements stimulating school internal accountability, which in its turn was seen as a positive condition of student achievement. An alternative interpretation could be that schools diminish internal evaluation if they have to exert themselves for abiding to external accountability requirements (leading to a negative correlation between accountability and student performance, Scheerens et al. 2015).

Autonomy at the lower aggregation levels, schools, classrooms and students increase the complexity of the model considerably. Potential for higher level control diminishes when controlled systems have autonomy, including the possibility to disagree on the overall goals of the system. Mintzberg's (1979) conception of the school as a professional bureaucracy is very adequate in explaining the limitations of direct control measures for this organization type. He describes professional bureaucracies as being critical of technological innovation and resistant against external evaluations. Professionalization and training represent the most effective coordination measure for this organization type. Weicks's construct of loose coupling goes beyond professional autonomy in describing the fragmented nature of the functioning of schools as organizations (Weick 1976). Seen from the perspective of higher policy measures influencing and controlling the behaviour of

lower level parts of the education system, autonomy impels the recognition of implementation issues. The respective literatures on policy implementation and planned change in education indicate that implementation is more often a process of "mutual adaptation", rather than straightforward "fidelity" to system-level inputs cf. Ball et al. (2012), Elmore (2000), Miles (1998) and Mitchell and Sackney (2000).

In Chap. 11 implementation problems, unspecified causal mediation, strong composition and ecological effects and small elasticities[2] were mentioned as potential conditions of "ineffectiveness". All of these phenomena are credible as explanations for the mostly low effect sizes, general stability (lack of change) of educational systems over time and the impact of background conditions being generally higher than that of malleable process conditions. The consideration of teacher autonomy enforces the relevance of teacher training and teacher characteristics, and the presence of important composition effects draws the attention to deliberate selection as a relevant control measure. With respect to selection, particularly early selection into a specific category of secondary education, the research results with respect to tracked and comprehensive systems are quite telling, as it came out as the one malleable system-level characteristic that was consistently associated with achievement levels and equity indicators.

The Overridingly Strong Influence of the Rational Planning Model

Educational effectiveness thinking is intrinsically associated with the rationality paradigm. Practically all of the conceptual models and middle range theories of educational effectiveness, reviewed in Chap. 11, are based on particular interpretations of the rationality paradigm. Those approaches that emphasize a harmonious connection of organizational and technical core activities of schools (like the Creemers and Kyriakides model) or stress a fitting connection of organizational and technological arrangements with the external environment (like the Quinn and Rohrbauch competing values model, Quinn and Rohrbauch 1983) are built on the premises of contingency theory. Contingency theory can be seen as an even more demanding version of the rationality paradigm than synoptic planning, as it includes an interaction effect in its optimization formula: not just optimal means to reach a particular goal, but optimal means given certain "other" conditions. In the image of schools as High Reliability Organizations, the organizational structuring component of the rationality paradigm is the core issue.

Educational effectiveness research itself can be seen as having a place in evidence-based, rational policy making, in search for those inputs and means that optimize the attainment of educational goals.

[2]Degree to which a dependent variable (such as expenditure) changes in response to a change in an associated independent variable (such as income).

As a prescriptive meta-theory on "good" policy making, management and teaching, the rationality paradigm, and its more partial interpretations like cybernetics and public choice theory, offers a normative framework to which empirical research results can be compared. Do the factors that are proven to have a positive association with educational performance work because they manifest some of the key components of rationality (like clear goal statements, proactive structuring, monitoring and feedback)? And, the other side of the coin: are failing approaches to be interpreted as failures to work rationally?

As noted in Chap. 11 each of the major interpretations of the rationality paradigm has a central mechanism. These mechanisms, *proactive structuring* for synoptic planning (also sometimes described as the pure rationality model), *fit* for contingency theory, *retroactive planning* for cybernetics and *market mechanism* for public choice theory, can be used as a basis to categorize major levers and independent variables in educational effectiveness (see Table 11.2). The empirical evidence for the effectiveness of the variables that represent each of these four mechanisms will be briefly reviewed in the sections below.

Proactive Structuring

Although attempts at top-down synoptic planning as in the X-year plans of communist regimes are generally refuted in most countries, nowadays, the principle of proactive structuring is "everywhere" in ideas on educational reform and school improvement. At system level it is evident in the work on national core-curricula, and national efforts to assure the alignment in educational systems between curriculum, assessment and instruction (see OECD 2010b). At school level the concept of High Reliability Organizations, although not yet applied on a large scale, is a case in point, and at classroom level, goal setting, systematic planning of didactic actions and use of monitoring and feedback frequently function as the backbone of instructional improvement programmes (cf. Locke and Latham 2002; Visscher and Ehren 2011).

The supporting evidence for the approach of proactive structuring is relatively strong. Combinations of opportunity to learn, time management, and clear and structured teaching showed the highest effect sizes in our review of meta-analyses, although in absolute terms the effect sizes were modest.

As possibly negative effects of some of the approaches associated with proactive structuring, goal displacement and bureaucratic process control (checking on standard operating procedures) were mentioned.

Internal and External "Fit"

Contingency theory emphasizes good internal fit of different organizational and technology components and external fit to environmental conditions.

Comprehensive school reform programmes offer an aligned combination of using externally developed, evidence-based curriculum materials, development of school leadership, professional development of teachers, community involvement as well as monitoring and assessment. This approach can be seen as an example of stimulating internal fit between several dimensions of schooling. Results of meta-analyses are positive but small (average effect size of $d = 0.15$, Borman et al. 2003; Hattie 2009, 215). In Chap. 3 the Quinn and Rohrbauch model of organizational leadership was mentioned as an example of fitting leadership approaches to external conditions. Individual studies show large differences in effect sizes. Griffith (2003), found effect sizes of the order of 0.40, Ten Bruggencate et al. (2013) near-zero effects.

Retroactive Planning

Evaluation is usually presented as the last phase in a proactive planning cycle, but it can also be organized as the first step in a planning cycle, hence the term "retroactive planning" (Scheerens et al. 2003, Chap. 5). Accountability policies, school self-evaluation and the use of pupil monitoring systems at school level can all be analysed for their effectiveness enhancing potential. Despite the strong logic, and the way these evaluative approaches can be applied in government programmes that seek to combine autonomy over process and control on outcomes, the research evidence at system, school and classroom level is mixed, at best.

Implementation failures and resistance from schools and teachers seem to constrain the potential of these approaches in important ways. External school evaluation is the arena where the professional autonomy of teachers and schools clashes with externally induced rational techniques, as convincingly predicted by Mintzberg (1979) in his model of the school as a professional bureaucracy. When schools and teachers agree to cooperate, lack of capacity to work with data, and to systematically plan action alternatives, may manifest itself as another kind of difficulty (Visscher and Ehren 2011). Finally, evaluation and monitoring activities enlarge the workload of teachers and schools, sometimes requiring exaggerated additional administrative work.

Market Mechanisms in Schooling

Competition between schools, free school choice, private schooling and school autonomy all fail to show convincing positive effects on student performance. A basic reason is that free school choice, when it happens, does not necessarily stimulate schools to be more effective. Parents are not primarily guided in their school choice by assessments on school performance indicators, but often give predominance to other considerations, such as the geographical closeness of the school to the home, and specific eye-catching special features schools offer to attract students (Pannecoucke 2005). Levin argues that in actual practice schools are

rather conservative and not entrepreneurial in a way that improves education (Levin 2006). Evaluation results of the American Charter Schools created to benefit from autonomy and competition with public schools are rather disappointing (Miron et al. 2010). Micro-economic theory seems to do better in analysing ineffectiveness when discussing sources of off-task behaviour, "make work" and exaggerated managerial overhead.

Alternative Theories and Mechanisms

To a degree the alternative theories discussed in Chap. 11 differ from the rationality paradigm in the sense that they are descriptive, rather than prescriptive. As such they may be seen to provide images of how schools as organizations and educational systems at large are "really functioning". From the perspective of these "alternative theories" the reality of schooling is not seen from the normative perspective of goal attainment and improvement, but is approached in a less pre-structured, open descriptive way. Weick's theory on loose coupling shows a complete opposite to well-aligned organizational structures, and is better fit to analyse autonomy and how it manifests itself. In important ways loose coupling can be used as an explanation for why rational techniques do not work, or at least do not work as well as expected. Implementation failures can be seen as consequences of fragmented organizational structures, liberty for idiosyncratic interpretations and lack of coordination. Sub-optimal application of evaluation and feedback mechanisms as a consequence of immunization against external review, limits to organizational learning and the overall treatment of evaluations as alien elements in the normal routine of the school's functioning. More positive interpretations of loose coupling are: keeping coordination and leadership costs low, ensuring certain basic and robust informal agreements as instances of "weak coupling" and as functional to "unfreezing" phases of innovatory programmes.

Next to loose coupling, "emergence" and self-organization were presented as a second opposite to rational planning. In the complexity sciences dynamic interactions at the micro level of systems are seen as a basis for emerging rules and regularities. These interactions are seen to take place in the informal part of the organization, comparable to the deeper structures of the organization culture, such as described by Schein (1985); see Chap. 3 of this book. Entrance conditions of the micro-elements are considered as very important for the way the dynamic interactions and emerging structures are developing. This is why a connection was made with the importance of composition effects for the performance of educational systems. If the composition of the student population or the teaching staff of a school is seen as "given", their (important) implications could be seen as a product of spontaneous development and self- organization. This interpretation is somewhat forced, however, since composition can also be seen as the product of more or less managed selection, admission and matching procedures. As such, it draws attention to a different strategy for enhancing the performance of educational systems, set apart from the usual interpretation of effectiveness enhancing conditions, in terms of

organizational and instructional strategies associated with performance. For the sake of theoretical argument it should be added that as soon as composition is deliberately planned and managed, it should not be seen as an instance of self-organization.

All in all the alternative theories do better in explaining ineffectiveness of rational strategies than in offering viable alternatives for educational policy and management strategies. This, however, is to be seen as an important result, as knowledge and better understanding of the phenomena described under the headings of loose coupling and self-organization forewarns against exaggerated reformist optimism. Loose coupling, school cultures that reflect schools' "regulating themselves through their dominant rituals, symbols, and images, irrespective of the degree of external regulation" (Levin 2006, 28) and resistance to external change, are also mentioned in seminal contributions from the US: Sarason (1990), Meyer and Rowan (1977), Cuban (1990), Elmore (2000) and Levin (2006). All these contributions suggest that high expectations of dramatic progress in educational reform should be generally toned down.

The Implications for Educational Policy and Practice

Relatively low effect sizes of effectiveness enhancing conditions in empirical results, theoretically interpreted as "constrained" rationality, caused by the loosely coupled nature of educational systems, lead to a prudent outlook on educational reform and improvement. Still, the results that were presented provide indications of the pros and cons of different improvement strategies that may have stronger or weaker claims for success. Before taking a closer look at alternative strategies a brief excursion will be made to the fields of knowledge utilization applied to educational effectiveness and approaches to school improvement.

Utilization of the Educational Effectiveness Knowledge Base[3]

In their paper "Getting lost in translation" Harris et al. (2013) discuss the "International Engagement of Practitioners and Policy Makers with the Educational Effectiveness Research Base" (Scheerens 2014). As a key source for the state of the art of the research base they refer to review papers on educational effectiveness, teaching effectiveness and system reform and school improvement by Reynolds et al. (2014), Muijs et al. (2014) and Hopkins et al. (2014). In their analyses

[3]This section is based on Scheerens (2014) The ripples and the waves of educational effectiveness research., published in *School Leadership and Management,* which is a commentary of the paper by Harris et al. (2013). Permission to use this material was granted by the publisher of the original source, Taylor and Frances, http://www.tandfonline.com/.

Harris et al. (2013) appear to assume that the knowledge base on educational effectiveness is solid, at least they do not raise critical questions about it. They note a sub-optimal engagement with and use of the available knowledge by educational practitioners and policy makers, and attribute this to diverse, mostly communicative weaknesses at the supply side (researchers) and demand side (practitioners and policy makers). They mention a long list of supposed weaknesses at the supply side of the research, such as: academic language of reports, sheer volume of research, insufficient effort from researchers to make their findings accessible. The article also mentions demand side problems and reasons why politicians and practitioners insufficiently engage with the educational effectiveness research base, offering ways of improving this state of affairs. In their analyses of supply side explanations for underutilization the authors are more critical of school effectiveness and teacher effectiveness research than about school improvement research. This may not come as a surprise given the intrinsic stronger preoccupation of improvement research with issues of transmission and implementation of knowledge. Features of educational effectiveness research on which the authors are particularly critical are: a narrow orientation with respect to basics, behaviouristic models in teaching research, check-lists of factors that are supposedly universally valid, lack of theoretical explanation, the statistical language of research reports and a backward, rather than a forward looking approach in the research orientation (concentration on what worked in the past instead of what will work in the future).

Although some of these critical comments are certainly justified the criticism seems to focalize exclusively on quantitative research, turning a blind eye on the weaknesses of qualitative studies, which may have easier co-involvement from practitioners, but lack epistemological validity and generalizability, and present serious problems as it comes to creating a synthetic knowledge base.

Research utilization can be modelled as a sequence of (1) actual research and knowledge production, (2) application and implementation of research results and (3) actual use, in the sense of sustained application or institutionalization of improved practice or policy. To return once more to the difference between educational effectiveness research and improvement research: the educational effectiveness researcher finishes his or her job with phase 1, whereas for the improvement researchers things start to become most interesting at the initiation of phase 2.

The utilization model gives a first hint of the complexity of research utilization. Remaining with the metaphor of something getting lost, things may go wrong from phase one: how solid is our educational effectiveness knowledge base? In the later phases of implementation and use completely different issues come into play, involving communication and the receptivity of the users, in the most general sense. The literature on research utilization, particularly also about the use of evaluation research by policy makers, highlights the intricacies of these processes. Weiss (1980) states for example that use is more often “conceptual” rather than “instrumental” and a slow process of information “trickling through”. To cite Patton’s (1978, p. 31) phrase: “research impacts in ripples, not in waves”. The overall assessment that is made in “Lost in translation” is that “many schools and

educational systems are not making enough use of the material and insights" from the educational effectiveness and school improvement knowledge base.

What was written about the educational effectiveness knowledge base in Parts 1 and 2 of this book shows consistency in the delineation of "factors that work", but considerable divergence in the effect sizes of the selected factors. This makes the findings of educational effectiveness research debatable and this might explain some of the underutilization that is suggested by Harris et al. (2013).

The most challenging part of the paper by Harris et al. is the final section where the authors state that instead of the incremental progress in compartmentalized separate fields, the way forward for educational effectiveness research lies in a "more tightly integrated and fully iterative field" as a basis for "the influence on policy makers and practitioners that it deserves". The authors speak of a new paradigm with the following main ingredients:

- a multi-level structuring of the field, supposedly integrating system, school and teacher effectiveness;
- a more proactive approach in influencing policy makers, where the recent McKinsey reports on well-functioning educational systems are set as an example (cf. Mourshed et al. 2010);
- more powerful combination of different research methodologies, which seems to be a call for "mixed methods" research;
- inverting the dominant model of research informed policy and practice, to a more interactionist approach in which researchers, politicians and practitioners "generate a true community of expertise".

I would agree that there is a lot of promise in following up on the first three points. Multi-level integration of types of educational effectiveness research is a most relevant challenge for a more model- and theory-driven research practice. Not all is gold that glitters in the "McKinsey approach" (see, for example Coffield 2012), but the research world can learn from its success in getting the attention of policy makers across the world. Mixed methods research is an interesting idea, with a lot of potential, but requiring solutions for credible and verifiable ways of combining different kinds of data (Scheerens 2012a, b).

The fourth claim recommends that educational effectiveness research should follow the model of mixed professional learning communities more frequently, to a degree that it should even become the predominant approach. Although I can see that there is a place for participatory approaches, development research or design-oriented research in educational effectiveness and improvement, I do not follow the authors in their claim that this approach should become predominant. The beauty of the interactionist paradigm is that it would seek to solve problems of knowledge production, implementation and use all in one blow. It fails, however, to recognize that trade-offs exist between organizing the field to answer research questions in a rigorous way (e.g. by means of randomized field trials) and full participation. As researchers we need the whole spectrum of available research methodology to strengthen the educational effectiveness knowledge base and as

facilitators of school improvement the whole gamma from externally programmed to internally co-constructed implementation.

Implementation of educational effectiveness research findings can also be seen against the background of introducing rational procedures and instruments in organizational structures, characterized by loose coupling and self-organization. A strong example of how two make these two worlds meet is Elmore's analyses on the introduction of standard-based accountability, supported by distributive, instructional leadership (Elmore 2000). In more general terms the literature on educational change and school improvements has offered different strategies to make the best of development and implementation along the lines of the educational effectiveness knowledge base; these will be addressed in the subsequent sections.

Strategies of Systemic Reform and School Improvement

Terms like "systemic reform" and "restructuring" are used to refer to changes in the institutional and organizational infrastructure of a country's education system.

New institutional economics, cf. North (1990) emphasizes the importance of institutions, where institution are "the rules of the game" in the form of basic legislative arrangements. "Marriage" and property rights are frequently mentioned examples of institutions. In developing countries quality assurance or quality improvement might start with an analysis of the institutionalization of basic arrangements in the education system, like: fixed teacher salaries; responsibilities of head teachers, official working time, timetables, etc.

Organizational infrastructure might have to be scrutinized as well. The "organizational capacity" of the country's Ministry of Education, for example, might be analysed and found to be in need of improvement (Orbach 1998).

Questions about organizational capacity of an educational system first of all regard the issue of whether core functions have an "organizational home" in the system. For example, initiating a national assessment is the heavier task when there is no organization that has specialized in the development of educational achievement tests in the country. The same applies when external supervision of schools is considered at a fairly large scale and the country has no educational inspectorate.

Further criteria in determining the organizational capacity concern the well-functioning of organizations in terms of effective leadership, ability to mobilize financial, material and human resources and appropriate work practices (ibid.).

The division of decision-making authority across the administrative levels of the educational system has both institutional and organizational significance.

In Western countries "restructuring" and systemic reform are usually focused at decentralization of decision-making authority and creating arrangements for accountability. Sometimes these two major dimensions are combined in certain patterns or arrangements that gear decentralization to accountability arrangements.

So-called "Performance-based approaches to large-scale reform" (Leithwood et al. 1999) form a case in point.

School improvement as a field of academic study is seen as a specific branch of the study on educational change. In some applications it is explicitly related to the school effectiveness knowledge base, and, in still other applications, its insights are also combined and integrated in perspectives on systemic reform.

Matthew Miles' overview of the development of this field of study, in the period between the mid-1950s and the mid-1990s provides the flavour of what the field represents. Miles discusses ten consecutive school change strategies:

1. *Training for group skills*; i.e. teaching school people fundamental skills of group behaviour. Such skills were considered to be of key importance for developing cooperation, "process analysis" and self-reflection for school teams and also as a subject to be addressed in classroom teaching.
2. *Innovation, diffusion and adoption*; which came to the fore during the 60s. This was the period of programmed instruction and the idea of "teacher proof" curriculum materials, i.e. curriculum materials and teaching methods that were specified to such a degree that bad teaching could not spoil the deliverance to students. And even when it was recognized that teaching material required an "interpretation" by teachers, for example in adapting to local conditions, the criterion for successful implementation was coined in terms of the "fidelity" to the externally determined "script". Educational change experts like Miles, at first, thought of temporary systems, like task forces, that were more actively involved in interpretation, and later on developed concepts on more active adaptation and re-creation of externally induced innovations.
3. *Organizational self-renewal.* Following developments in industry that came under the heading of "organizational development", during the late 60s the school as an organization was increasingly seen as the object and the agent of change. The aim was "to induce organisational self-renewal through tactics of training, process consultation, data feedback, problem-solving and structural change" (Miles 1998, p. 48).
4. *Knowledge transfer.* In this area the expectations of the Research Development and Dissemination (RDD) strategies were challenged, again (as in point 2) emphasizing active reconstruction at school level of the knowledge that was offered from outside. Capacity building was seen as a necessary prerequisite of good knowledge transfer.
5. *Creation of new schools.* In this section Miles writes about the phenomenon that many new and alternative school projects came into being in the 1960s and 70s in the USA. He concludes that by analysing some of these, he learned that "good new schools can be created, but that the task is very demanding, more complex than expected, and requires assistance and political protection" (ibid., 50).
6. *Supported implementation.* In the late 70s the "passive" idea of adoption of externally induced change had been abandoned, and instead, implementation was being seen as a longer term process of "adaptation". Adaptation required

that schools develop coherence and meaning to external change initiatives. Based on his experiences of assisting some large projects that recognized this implementation perspective Miles concluded that "continued assistance" throughout the implementation process was of major importance.

7. *Leading and managing local reform.* According to Miles during the 80s there were a lot of local initiates "many of them pushing hard on effective schools and effective teaching programs". On the basis of studying successful projects, Miles and his colleagues identified characteristics of success local reform projects. He summarizes as follows: "The ideas of vision-building, pressure and initiative taking, and assistance have already been outlined. The idea of *empowerment* is an extension of the concept of *legitimacy for planning and action,* indicating in sharper terms that we found reform success closely associated with the presence of a cross-role planning team with clear decision power over change-related matters (such as project budgets, staff development, staffing patterns, and related time)". He goes on to say that he found three variables that were tied to successful local reform. The notion was that the planning style was "evolutionary" rather than "architectural". He describes evolutionary planning as "a journey in the service of an evolving, increasingly shared vision". Secondly he found that successful schools were good at resourcing and problem coping (the slogan: "problems are our friends").
8. *Training of change agents.* As concluded earlier, despite the importance of local initiative, school change is usually in need of external support and facilitation, according to Miles, where the support is primarily focused at the process of change. In this context he identified two major characteristics of successful support. "*Developing trust and rapport.* A great deal seems to depend on a change agents' ability to develop a strong, supportive, contractually clear relationship with specific "clients"—groups and individuals involving in change efforts". The second characteristic of successful support is *organizational diagnosis*, a data-driven "understanding" of schools as organizations.

The last two strategies that Miles mentions are in fact about integrating school change in system- wide reform initiatives: "managing systemic reform" and "restructuring schools". Murphy (1993) states that "restructuring" in the USA usually had four main strategies for reorganizing education: providing choice and voice for parents, school-based management, teacher empowerment, and teaching for understanding. The latter refers to a constructivist orientation to teaching and learning.

Taking into consideration other seminal contributions to the conceptualisation of school improvement, as those by Fullan et al. published in the "International Handbook of Educational Change" (1998) edited by Hargreaves et al., the following can be seen as the key principles of this orientation to educational change:

(a) The school is the focus of educational change. This means that schools should be analysed as organizations, seen in their local contexts and harbouring the major agents of change, namely teachers.

(b) A strong emphasis on the process dimension of educational change.
(c) The importance of school-based "implementation" in the sense of active adaptation or "co-invention" of externally induced changes.
(d) A human relations approach to educational change influenced by group dynamics and the idea of teacher "empowerment", capacity building and overcoming professional isolation of teachers. The "counselling" approach of external change facilitators perhaps also fits in this tradition.
(e) An evolutionary "bottom up" view on educational planning and curriculum development.

Within the scientific community active in this field quite a range of emphases can be discerned. These vary from authors like Mitchell and Sackney (2000), who provide a post-modernist view on school improvement and are strongly opposed to accountability and other "mechanistic" approaches, to authors like Reynolds and Hopkins, who relate school improvement to school effectiveness research in emphasizing learning and learning outcomes. Still other contributions (e.g. Leithwood et al. 1999; Hopkins 2002) integrate school improvement approaches and conceptualisations of systemic reform, where key characteristics are rigorous accountability requirements from the state or district level combined with support networks and capacity building at the school level.

A major breakthrough in this field is the work of Slavin, who proposed a "third" way, in addition to the school improvement approach and systemic reform (Slavin 1996, 1998). Slavin refers to the school improvement approach as described in the above as "organisational development models". "Perhaps the dominant approach to school-by-school reform is models built around well-established principles of organization development, in which school staff are engaged in an extended process of formulating a vision, identifying resources (such as external assistance, professional development, and instructional materials) to help the school towards its vision, and often locating "critical friends" to help the school evaluate and continually refine its approaches". Of this approach Slavin says that it is time-consuming and expensive. Moreover, he claims that it is only effective for schools that already have a strong capacity for change. "Such schools are ones in which staff is cohesive, excited about teaching, led by a visionary leader willing to involve the entire staff in decisions, and broadly aware of research trends and ideas being implemented elsewhere" (p. 1303). Such schools he describes as "seed" schools. A second category of schools Slavin describes as schools that would like to do a better job, but do not perceive the need for the capability to develop new curricula. According to his categorisation these are schools with good relations among staff and leadership, a positive orientation toward change, and some degree of stability in the school and its district. Finally, as a third category, he refers to schools "in which even the most heroic attempts at reform are doomed to failure. Trying to implement change in such schools is like trying to build a structure out of sand" (ibid., 1303). Accordingly he refers to these schools as "sand" schools.

School improvement of the organizational development kind (as we have seen the predominant perspective on school improvement) is considered only feasible in

"seed schools", which he estimates at 5 % of all schools in the USA. Sand schools, also about 5 % of all schools would require fundamental changes before they can support any type of school change. The overall majority of schools, according to Slavin, are the brick-schools and they could most efficiently benefit from what he calls *comprehensive reform models.* His "Success for All" programme is an example. Comprehensive reform models provide schools with specific student materials, teachers' manuals, focused professional development, and relatively prescribed patterns of staffing, school governance, internal and external assessment and other features of the school organization. It should be marked that "Success for All" is one of the few improvement projects that has been thoroughly empirically evaluated and has shown to be successful (Slavin 1996; Scheerens and Bosker 1997). Similar successes have been reported by Stringfield et al. (2011) presenting the idea of schools as "high reliability organizations".

It is interesting to note that Slavin's conception (and also its actual realization in "Success for All") of Comprehensive Reform Models, seems to have returned full circle to the point where, according to Miles, the school improvement movement started its human relations/implementation approach in the 1950s. Namely, the discussion on the applicability of externally developed pre-structured innovation programmes and curriculum material. The fact that there is clear evidence that this approach works is revolutionary, and puts a question mark behind the efficiency of 40 years of educational innovation based on the less directive, bottom-up, social psychological, organizational development approach to school improvement. The question of efficiency is rarely raised from within this tradition, so eloquently described in Miles' ten strategies for school change.

A critical issue in combining systemic reform elements (particularly standard-based accountability) with decentralization and autonomous adaptation at local level is whether these approaches should be seen as complementary and functional or as working against one another. Muijs and Reynolds write somewhat polemic about contrasting what they describe as "the ownership paradigm" and the externally supported approach (compare Slavin's approach, outlined in the above). Elmore's essay on "new structures for school leadership" (Elmore 2000) takes a very interesting position in this debate. He is quite emphatic about two points: the loosely coupled nature of the way American schools operate, causing ineffectiveness, and the introduction of standard-based accountability as the mechanism that has the potential to bring about improvement. He describes distributed instructionally oriented leadership as the mechanism at school and above school level that should be capable of overcoming loose coupling and fostering commitment for improvement. In the more recent developments to implement a common core curriculum in the United States, alignment between accountability standards and content selection is likewise introduced as a potentially effective lever for improvement (Schmidt et al. 2010).

What should be added to these considerations is that autonomy and doing away with externally induced technocratic and "mechanistic" influences has considerable support in the education sector. In a country like the Netherlands, which already has

very high school autonomy, at regular intervals new schools are founded that decide to do away with textbooks, learning programmes and timetables and propagate completely autonomous and self-regulated learning. Usually such initiatives perish in a couple of years, only to give place to new initiatives based on the same ideas. Similarly, school improvement initiatives that call for bottom-up innovation and empowerment of teachers frequently choose teacher cooperation as the main strategy, perhaps not so much for it being considered as the most effective strategy, but because it is less controversial and considered more motivating to teachers. For teacher cooperation to be effective it should be work related and achievement oriented (cf. Elmore 2000; Lomos et al. 2011). And finally, side effects of accountability policies should neither be overlooked (cf. Koretz 2005) nor overrated (Hanushek and Raymond 2006).

Levers for Improvement

The Substantive Focus of School-Level Strategies

The school and educational effectiveness knowledge base provide an instrumental orientation to school improvement, meaning that enhancing identified school factors is expected to lead to better student performance. In very broad terms the variables identified in educational effectiveness have to do with the technology of the curriculum (as intended and implemented) and with facets of the organizational climate. In this way one could say that a first broad orientation to school improvement could be labelled as the *curriculum and climate emphasis*.

However, it should be noted that schools can choose alternative orientations. A second strategy might be labelled the *teacher recruitment and professional development strategy*. According to this strategy most of a school's energy to improve should be focused on teacher issues, including human resources management and capacity building. In the third place, schools could capitalize on *matching and grouping issues*. Matching could be both externally oriented, towards the local community, towards higher administrative levels, other schools and to parents, and internally oriented in grouping of students in classrooms and learning groups and assigning teachers to these groups of students.

As noted in the above, applying the knowledge base of educational effectiveness research is closest to the curriculum and climate orientation. The general factors that have been discussed and rank-ordered in previous sections are all candidates to be stimulated. More minute and detailed descriptions of these variables were described in Part I of this book, other relevant references are: Marsano (2003), Scheerens et al. (2007) and Hattie and Alderman (2012). On the level of strategy choice a more synthetic description of the key factors is considered helpful. The following alternative emphases within the curriculum and climate orientation are distinguished:

(a) *Exposure to educational content*. This could be seen as a composite of opportunity to learn and instruction time. It expresses the curricular focus and duration of exposure in school curricula and teaching.
(b) *Evaluation, monitoring and feedback provisions*. Evaluation and feedback can be seen as driving improvement at school and classroom level. Implied facets are clarity of purpose through standards, examination syllabi, etc., verification of what students have learned, identification of strengths and weaknesses in content and skills that are mastered, feeding back and diagnosis of outcome patterns, systematic consideration of remedial strategies and setting concrete goals for improvement at student, classroom and school level, in cooperation with other teachers, school principals and eventual support staff. This latter characteristic could make evaluation/feedback/systematic corrective action the core of task-related professional development and teacher cooperation.
(c) *Managing the school climate*. This involves diverse facets like creating a safe atmosphere, positive interactions as well as fostering high expectations and pressure to achieve.
(d) *Managing the teaching and learning program*. Repeated studies, in which more behaviouristics approaches like "direct teaching" were compared to constructivist approaches and where no significant differences in student achievement were found, have inspired analysts to propose more general underlying constructs. One example is the construct of "cognitive activation" (Klieme 2012), discussed earlier. Another example is the term "focused teaching" coined by Louis et al. Hattie (2009) proposes "active teaching" as an overall construct. Careful attention to lesson planning, variation in structure and independence in learning assignments and keeping students engaged seem to be the core issues in these constructs.
(e) *Meta-control* as the overriding leadership approach. Meta-control is a concept from control theory, and literary means "control of controllers". Applied to school leadership this concept emphasizes the notion that schools are professional organizations, with teachers as semi-autonomous professionals. Teachers may be metaphorically seen as the prime "managers" of teaching and learning at school. A school leader as a meta-controller is not a laissez-faire leader, but one who sets clear targets, facilitates and monitors the primary process of schooling from a distance (Scheerens 2012a, b).

These five strategic angles to enhancing school effectiveness can be seen as having certain connections. Exposure and evaluation/feedback have a common element in educational objectives and learning standards. Alignment of what is taught and what is tested is the key issue of opportunity to learn. High expectations and pressure to achieve, as facets of the school climate, likewise need a substantive focus in the form of objectives, standards, assessment instruments and feedback. The educational content dimension, perhaps indicated as the implemented school curriculum, is a core dimension of the teaching and learning programme, next to ideas on transmission that are central in concepts like cognitive activation. Managing all of these strategies, as well as their connections, is the task of school

leadership as meta-control. Integration of these angles to school improvement, inspired by the educational knowledge base is close to the approach of Comprehensive School Reform, e.g. Borman et al. (2003).

How System-Level Policies Could Foster These School-Level Strategies

System-level policies and structural characteristics of educational systems can be seen as preconditions or constraints of school-level improvement policies, to which schools need to adapt. More analytically one could ask which system-level conditions could be seen as supportive of effective schools and effective school improvement. A third, more "neutral" approach might be to just establish where there are matches between the major system-level reform dimensions and structural conditions, as discussed earlier, and the school-level improvement strategies.

Following this third approach would favour accountability policies as the best matching system-level arrangement for the curriculum and climate orientation to effective school improvement. Accountability policies touch directly on core facets of school functioning, like performance standards, achievement orientation, and perhaps also on the "internal accountability" of schools (see the earlier discussion on the work of Carnoy et al. 2003). As accountability policies are almost inevitably associated with a degree of centralism in the curriculum, this would emphasize the connection with content exposure and opportunity to learn at school level.

Other system-level policies and structural arrangements are more closely associated with other orientations to school improvement than the curriculum and climate orientation. Enhanced school autonomy as well as strong teacher policies appeal more to teacher recruitment and professional development and to organization development. Choice and market mechanisms, as well as tracked versus comprehensive school systems, are more associated with admittance, selection and grouping processes at school level.

A Final Note on the *Process* of School Improvement

Among the classic change strategies proposed by Bennis et al. (1969), considering school improvement on the basis of the educational effectiveness knowledge base is a clear example of the family of rational empirical strategies. Rational empirical strategies assume a neutral position between "top-down" and "bottom-up" processes, and innovation strategies that embody these extremes: power coercive strategies on the one hand, and normative re-educative strategies on the other. In actual practice the dominant approach in school improvement, starting from the social psychological approach of Matthew Miles, has been bottom-up development. See the section on school improvement in Chap. 2. More recent developments like Comprehensive School Reform programmes, as well as calls for "evidence based" education policy oppose this dominant trend, at least to the degree that room for more external input to school improvement is implied. It might be argued that

external input in the form of assessment instruments, guidelines to interpret test scores as well as aligned syllabi and textbooks, are not to be seen as coercive prescriptions, but rather as inputs that always allow interpretation and adaptation by professionally autonomous teachers. Yet, in case of high school autonomy "evidence based" external inputs may evoke strong opposition, as will be illustrated in the case study on the Netherlands, presented in the next chapter.

Apart from strategic considerations and implementation issues, the core critical question of this review should be more substantively oriented: do we know enough to provide a strong evidence-based input to the practice of school improvement? The effect sizes of the main variables in the realm of what has been called the curriculum and climate orientation to school improvement are to be seen as small, when compared to general standards (Cohen 1977) and medium when they are compared to more arbitrarily standards of "educational significance" (see Chap. 8). It could be argued that the effect sizes reached with comprehensive school reform programmes, of the order of a *d*-coefficient of 0.15, are about what we can obtain by optimizing curriculum and climate at school level. Perhaps some success stories of national reform and improvement as evident from PISA are slightly more optimistic, like the progress made by Germany after the "PISA shock" in 2000, and the improvement of Polish results after integrating vocational and general secondary school tracks.

The alternative orientations, the teacher recruitment and professional development perspective and the matching and grouping orientation might offer additional and maybe even stronger effects. The evidence for this is more coincidental. If one looks at the excellent performance of Finland, this could be seen as strong evidence for a teacher centred orientation (Sahlberg 2011), but, although fundamental, improving initial training and professional development of teachers are at best very slow and time-consuming levers of educational improvement. The potential of the matching and grouping orientation might be inferred from the strong impact of school composition (in terms of the school average socio-economic status of the students' home background), as established from, among others, PISA data. Yet, optimizing school composition, through selective student intake policies, would usually be considered as stimulating "excellence" at the cost of equity in schooling.

The way the success of schools and students depends on the socio-economic background of the students and the school and classroom composition give reason for a prudent stance to the degree of malleability in education by means of what was indicated as the curriculum and climate-oriented strategy.

The Connection Between Educational Effectiveness, School Improvement and Systemic Reform Reconsidered

School effectiveness research and school improvement were closely connected, if not simply one, in the early days of the effective schools movement (Edmonds 1979). The factors that distinguished effective schools from ineffective schools, like

strong leadership, frequent evaluations, an orderly climate and high expectations, were seen as levers for improvement. In this way school improvement could be seen as putting the results of school effectiveness research into practice. At the same time, research depended on practical programmes to improve schooling, with a particular emphasis on schools that served high proportions of disadvantaged learners (cf. the work of Teddlie and Stringfield 1984 with respect to the Lousiana school improvement programme, and the study by Mortimore et al. 1988 in the inner city of London). However, among academics working in this field, and, since the 1980s united in the International Congress on School Effectiveness and School Improvement (ICSEI), a gradual specialization between effectiveness research and studying improvement processes developed. As described in the earlier sections, school improvement developers and analysts became particularly focused on change processes and implementation issues. Where school effectiveness researchers used predominantly quantitative, survey type of methods, school improvement researchers frequently applied qualitative research methods and case studies. More recently a new blending of the two orientations (school effectiveness research and school improvement analysis) can be discerned. Some of these have a predominant school effectiveness orientation but have added a dynamic element (in studying development over time) as well as a comprehensive outlook on the inclusion of a broad spectrum of effectiveness enhancing conditions. Other study-based models in the school improvement tradition enlarge strategic facets of change, use educational effectiveness research only in a rather global way and include normative and motivational issues in visions on educational change. Some of these more recent developments in school improvement thinking appear to go beyond the early symbiosis of school effectiveness research and school improvement. This symbiosis was based on two principles: (1) improvement is ultimately expressed in measurable educational outcomes and (2) malleable conditions to improve, such as curriculum strategies and school organizational arrangements, are valued for their instrumental connection with educational outcomes. School improvement thinking may go beyond these principles by emphasizing educational outcomes that are difficult to measure, by intrinsically valuing certain methods and conditions (even when they do not have a proven association with outcomes), for the sake of common sense, "good practice", or "modernization" and, as it would seem, by sometimes pursuing change "for change sakes".

Finally, the availability of internationally comparative work on "systemic improvement" has added a third category of models, which include system-level structures and policies. In the sections below, illustrations of these three types of school improvement models will be given. They are important in the consideration of how educational effectiveness research outcomes find their way to educational practice. They also provide a basis for debate about the evidence-based nature of the various approaches.

Improvement Models as Direct Extensions of Educational Effectiveness Research

The conceptually most developed example in this category is the "Dynamic approach to school improvement", by Creemers and Kyriakides (2012). The improvement approach is directly based on their "Dynamic model of educational effectiveness", which has been referred to in earlier chapters (Creemers and Kyriakides 2008). Their dynamic model of educational effectiveness is a multi-level conceptual model, with a structure that is comparable to the models that were described in Part 1 of this book, and it is particularly explicit at the micro level of classroom teaching. Basically the dynamic approach has three main characteristics: (a) its content is based on the elaborate and comprehensive model of educational effectiveness, (b) its main "vehicles" to stimulate improvement are school self-evaluation on the one hand and teacher (in-service) training on the other, and (c), the structured approach to manage change is a rational planning model. The subsequent steps in the improvement approach are cited in Fig. 12.1

A second example is provided in the study by Day et al. (2009), titled "The impact of school leadership on pupil outcomes". It was a major, labour intensive study, on a relatively large sample of British primary and secondary schools, which improved their performance over a three-year period (2003–2005). From national census data all primary and secondary schools that "were effective and improving over a 3 year period" were identified. School performance data consisted of pupils' cognitive outcomes from the Key Stage national assessment and CSCE results. Change in average school performance in mathematics and reading over the three-year period in question (2003–2005) was used as the dependent variable. A broad set of leadership facets, as well as school and classroom conditions as intermediary variables, were included in the study, and were described by means of quantitative and qualitative methods. The authors presented a set of strong causal claims about the importance of school leadership in promoting school improvement and school effectiveness. They conclude by saying that their study "…provides

A) Establishing clarity and consensus about the general aims of school improvement by considering student learning as the main function of the schools
B) Establishing clarity and consensus about the aims of school improvement by addressing school factors that are able to influence learning and teaching
C) Collecting evaluation data and identifying priorities for improvement
D) Designing school improvement strategies and action plans by considering the available knowledge base about the factor(s) addressed
E) Monitoring the implementation of the improvement project through establishing formative evaluation mechanisms
F) Conduct a summative evaluation to measure the impact of the Dynamic Approach to School Improvement

Fig. 12.1 Subsequent steps of applying the dynamic approach to school improvement, cited from Creemers and Kyriakides (2012, p. 64)

evidence of an increasingly sophisticated understanding of the strong links that have traditionally been thought to exist between school leadership and student achievement" (p. 191; Further description and commentary on this study is provided in Scheerens 2012a, b). An interesting outcome of the study was that it suggested that different approaches were important, depending on the level of performance of schools, in other words depending on whether change was from low to average, or from average to high. This way of analysing progress results was followed up in the McKinsey study by Mourshed et al. (2010), which will be referred to further on, as an example of systemic improvement analysis.

Comprehensive school reform programmes and their evaluation present another important example based on improvement models that are direct extensions of educational effectiveness research. As they were described more fully in an earlier section of this chapter, they will not be further discussed here.

School Improvement as Strategic Change

Hargreaves and Harris (2005) used case study methods of successful organizations in various sectors, sports, business and education, to describe "leadership of performance beyond expectations" and the factors on which this depends.

The ingredients of exceptional leadership are summarized in 15 factors, some of which are cited below (Hargreaves and Harris 2005, p. 45):

F1: The fantastic dream

– A dream that is bolder and more challenging than a plan or even a vision

F3: The fight

– the impossible dream and improbable challenge produce a response of fight to overcome or circumvent obstacles

F4: Fundamental futures

– "these organizations create an inspiring future by connecting with the fundamental meaning of a classic and honorable past"

F5: Firm foundations

– Heroic leadership builds upon solid platforms created by predecessors: a very solid platform in terms of curriculum, behaviour and assessment that was a firm foundation

F8: Fast and fair tracking

– Organizations that perform above expectations mark, monitor and manage their progress towards success.

F10: High fidelity

these organizations have a deep faith in and faithfulness to their colleagues and the people they serve, and a higher purpose greater than any one of them

F11: Fraternity

– these organizations build and sustain four kinds of communities: communities of recruitment, communities of service to customers or clients, communities of professional practice within the current organization and communities of support that surround them

F15: Fusion leadership

– "these organizations invest in leadership and followership that raise and rally the performance of the organization by lifting up its members morally, emotionally and spiritually through a combination and progression of leadership styles and strategies"

Complex qualities are rendered in a narrative that appears to be aimed at being persuasive and missionary more than factual. Achieving excellence in education (school improvement almost appears as too modest a term) is described as a heroic enterprise: "The existing literature and, as we shall see, the evidence-base from 18 organizations in three sectors, show that leadership of performance beyond expectations is integrated, not fragmented. It is a fusion of complex qualities, not the expression of a single style. It is as much a mythological narrative journey as a logical set or sequence of elements. And, we will learn, it is an expression and embodiment of the paradoxes that make up extraordinary leadership practice everywhere: heroic and humble, charismatic and traditional, long term and short term, emotionally inspirational and technically effective, and remarkable and unremarkable." (ibid., 52).

There is no strong connection to the educational effectiveness knowledge base in these recommendations, with the one about "fast and fair tracking", coming closest to being an exception.

This example may be seen as an illustration of qualities in school improvement that go far beyond application of educational effectiveness research findings. There is much concern with motivating the field, leadership gets a strong emphasis, also as a way to ignition change processes, school autonomy is almost taken for granted, and "ownership" of the improvement activities by the school staff is underlined. Finally, the emphasis is on school organizational processes, leadership, teacher cooperation and "peer learning", and more recently also use of data and formative assessment, as a school-based approach. Effective teaching processes are not directly referred to, but indirectly approached by stressing professionalism of teachers, continuous professional development and teachers assessing each other's work. The tone is optimistic and idealistic, there appears to be little room for technical problems, persistent routines in teaching and resistance to change.

Models for Systemic Reform

In this section two national and one international example for research-based guidance for systemic reform will be briefly described. The national examples have a locus in the state of Victoria (Australia) and a reference to the province of Ontario in Canada. The example from an international comparative context is based on studies by the OECD and McKinsey.

Victoria

In their "School Work Force Development Strategy Paper for the Department of Education and Training in Victoria" the Boston Consulting Group (BCG 2003) sketches workforce recruiting guidelines and a four-phased school development programme for the State of Victoria.[4] The development programme starts off from a vision of four major trends in education over the next coming 10–20 years: continuing development of an '**authentic curriculum**', including the use of ICT. An authentic curriculum is described as "the idea that learning tasks should follow from students' everyday experiences and be relevant to their personal goals and backgrounds. More broadly, 'authentic curriculum' covers a range of pedagogical developments over the past decade, and in particular the shift from teaching to learning, from mandated learning tasks to self-directed learning experiences, and from a one-size-fits-all approach to a more tailored learning experience". In the second place they state that, "**school-community relationships** will become increasingly important in generating desired outcomes for both students and communities. This suggests the need—as a central part of the teaching role—to develop teamwork both within the school and with others in the community. The teaching workforce will need ongoing development in the area of community engagement to deal with this". Thirdly, they note a trend towards a more **flexible teaching workforce**, "with greater role specialization, more teaming and greater use of non-teacher specialists such as artists, scientists, academics and psychologists. To operate effectively in this environment, teachers will need to develop a more open, collaborative working style". As a fourth and final major trend they state that "an increasing focus on ***performance and outcomes*** requires better measures of student achievement and teacher effectiveness. In this environment, schools and teachers will need improved access to good data, as well as increased familiarity with and ability to interpret and use those data." (ibid., 11, 12). Recruitment strategies for school principals and teachers, capacity building and reward structures are the main policy inputs to drive the development in the desired

[4] http://www.google.nl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCYQFjAA&url=http%3A%2F%2Fwww.education.vic.gov.au%2FDocuments%2Fschool%2Fprincipals%2Fprofdev%2Fpcpdschworkforce.doc&ei=K_MkVbLrC4XAPM_FgcAJ&usg=AFQjCNHmcEsxqnEWEDPBbGMHIhKJHU_DVg.

direction (see the four trends). These inputs are to set in motion the development of a performance and development culture by means of induction programmes of teachers new to the profession, use of multiple sources of feedback on teacher effectiveness (including student outcome data), for individual teachers and teams of teachers, customized individual teacher development plans based on individual development needs, student learning school priorities and quality professional development to meet individual development needs. The "Self-Assessment Framework" for the schools looks at the extent (quantity of implementation) of these assets, but also to the development level of each. At the highest level, schools have evolved to "transformed schools", where all kinds of innovative approaches are tried out, and research is used to guide the development of new practices.

When comparing the approach for Victoria to research-based effectiveness enhancing mechanisms we see application of recruitment and capacity building of professionals (school leaders and teachers), the use of evaluation and feedback, and organizational arrangements of self-managing schools and teachers involved in task-related cooperation (receiving feedback from colleagues). The normative framework for the Victoria development programme is broader than attaining excellence in student performance, which is the core criterion in educational effectiveness research. The programme started out with a vision for the future with a modernization of schooling as described in the above; a more authentic curriculum, including ICT applications, more intensive school–community relationships, a more flexible teaching force and increased focus on performance and outcomes.

Hargreaves and Braun (2013) on Data-Driven Improvement and Accountability

In this contribution the authors take a critical stand against the mainstream of accountability policies in the United States: "In general, we find that over more than two decades, through accumulating statewide initiatives in DDIA (Data Driven Improvement and Accountability, JS) and then in the successive Federal initiatives of the No Child Left Behind Act and Race to the Top, DDIA in the U.S. has come to exert increasingly adverse effects on public education, because high-stakes and high-threat accountability, rather than improvement alone, or improvement and accountability together, have become the prime drivers of educational change. This, in turn, has exerted adverse and perverse effects on attempts to secure improvement in educational quality and equity. The result is that, in the U.S. Data-Driven Improvement and Accountability has often turned out to be Data Driven Accountability at the cost of authentic and sustainable improvement." (ibid., ii)

Ontario is mentioned as one of the best-case examples of how to combine improvement and accountability in a form that confirms to their standards of data-driven improvement and accountability. The authors provide a list of 12 recommendations to restructure accountability policies so that they become more improvement oriented. The most important of these recommendations are:

- To be less reductionist in the scope of what is measured for accountability and improvement purposes, to capture the full range of what a school values.
- Insistence on high quality data, valid and reliable measurements, including data on progress and growth.
- "Creating a culture in which data are valued in helping solve improvement questions and where there is reciprocal accountability between schools and the central office, in an environment characterized by trust rather than threat".

Their final recommendation in this list is to "Create a set of guiding and binding national standards for DDIA that encompass content standards for accuracy, reliability, stability and validity of DDIA instruments, especially standardized tests in relation to system learning goals; process standards for the leadership and conduct of professional learning communities and data teams and for the management of consequences; and context standards regarding entitlements to adequate training, resources and time to participate effectively in DDIA".

The combination of state level, school external accountability policies and practices on the one hand and collaborative practices at school has been elaborated in a more fundamental way by Elmore (2000), when he writes of school leadership practices that have high teacher involvement. A helpful construct that is missed in Hargreaves and Braun's analysis is the distinction between external and internal accountability. The concept of internal accountability is discussed in Carnoy et al. (2003) in a study in which the effectiveness of accountability regimes across American states is compared. One of their conclusions is that systems with high stakes external accountability regimes had the highest results in terms of student performance, and this could partly be attributed to schools developing evidently effective *internal* accountability regimes.

A rather worrying facet is the idea of "binding national standards … for the leadership and conduct of professional learning communities". This is an invitation to process control and enforcing "standard operation procedures" in an area that is relatively weak in its evidence on predicting educational performance.

Recommendations for Systemic Reform, Based on International Comparative Assessments

Data from international assessment studies were used for two reports that appeared in 2010, *Strong Performers and Successful Reformers in Education*. Lessons from PISA for the United States (OESO, December 2010), and *How the world's most improved school systems keep getting better* (Mourshed et al., McKinsey and Company 2010). Both reports looked at school systems that had improved their performance over a number of years and drew conclusions about structures and policies that might have caused this successful performance. The high-performing

education systems included in the OECD report were: Canada (Ontario), China (Hong Kong and Shanghai), Finland, Japan and Singapore. The examples of rapidly improving systems that were chosen were Brazil and Germany. In the OECD report the following set of characteristics associated with performance was proposed:

1. Developing a commitment to education and conviction that all students can achieve high levels.
2. Ambitious standards aligned with high stakes gateways and instructional systems (well-aligned testing system).
3. Developing more capacity at the point of delivery (high quality teachers and school leaders).
4. A work organization in which the teachers can employ their potential: management, accountability and knowledge management (flat organization, away from Tayloristic management, school autonomy).
5. Institutionalizing improved instructional practice (diagnostic skills, encyclopaedic repertoire, students enthralled, devoted to the improvement of their craft).
6. Aligning incentive structures and engaging stakeholders (high stakes examination systems, in collaboration with stakeholders).
7. External, professional and parent-oriented accountability.
8. Investing resources where they can make most of a difference (strong teachers aligned to weak students).
9. Balancing local responsibility with a capable centre with authority and capacity to act (state sets clear expectations).
10. Importance of work-based training (transition school work).
11. Coherence and alignment across levels, policies and practices, and sustained input (policy implementation).
12. An outwards outlook of the system (responsiveness).

In this list key issues are alignment across sub-systems (across levels and functional domains), evaluation and accountability measures, and improved instructional practice.

The Mc Kinsey report by Mourshed et al. (2010) describes the identification of 20 school systems that showed significant progress in performance since 1980, and the analysis of conditions that were seen as bringing about this performance gain. These school systems stood out positively when compared to a group of 40 other systems, which were either stable or declining in performance. The study's main interests were to identify the current performance level, to describe the set of interventions necessary that had made the desired improvement in student outcomes possible and to analyse the system's adaptation of the intervention cluster to the current context.

The school systems' performance levels were categorized into four broad groupings: poor, fair, good, great or excellent. Next, each system, with its interventions, was matched with a performance stage (poor to fair, fair to good, good to great and great to excellent). The study revealed intervention patterns that were seen as typical for each performance stage:

> Poor to fair: the interventions in this stage focus on supporting students in achieving the literacy and math basics: this requires providing scaffolding for low-skill teachers, fulfilling all basic student needs, and bringing all the schools in the system up to a minimum quality threshold.
>
> Fair to good: at this stage the interventions focus on consolidating the system foundations; this includes the production of high quality performance data, ensuring teacher and school accountability, and creating appropriate financing, organization structure, and pedagogy-models.
>
> Good to great: the interventions at this stage focus on ensuring teaching and school leadership regarded as a full-fledged profession; this requires putting in place the necessary practices and career paths to ensure the profession is as clearly defined as those in medicine and law.
>
> Great to excellent: the interventions of this stage move the locus of improvement from the centre to the schools themselves; the focus is on introducing peer-based learning through school-based and system-wide interaction, as well as supporting system-sponsored innovation and experimentation. (ibid., 20)

A common set of interventions was identified that operated across performance levels, in the sense that they occurred at each level, but with different manifestations. These core interventions were: an appropriate reward and remunerations structure for teachers and principals, building technical skills of teachers and principals, assessing students, establishing data systems, and facilitating improvement through the introduction of policy documents and education laws. Another interesting thesis that was developed in this study was the notion that the overall state-level improvement policies had different characteristics, depending on the performance stage. "Systems on the journey from poor to fair, in general characterized by less skilled educators, tightly control teaching and learning processes from the center because minimizing variation across classrooms and schools is the core driver of performance improvement at this level. In contrast, the systems moving from good to great, characterized by more highly skilled educators, provide only loose guidelines on teaching and learning processes because peer-led creativity and innovation inside schools becomes the core driver for raising performance at this level." (ibid., 20)

The study proposes a few other strategic facets of successful reform: thoughtful selection of intervention menus by the centre, followed up by a process of implementation characterized by fidelity, attention for the ignition of improvement programmes, and for sustaining them, after implementation. The first strategic facet (integrity of intervention menus and implementation) seems to express that successful reform is to be seen as orchestrated and planned. It is in line with the emphases on "alignment" expressed in several of the recommendations from the OECD report, cited in the above (OECD 2010). About "getting started" or ignition of a reform and improvement programme the McKinsey study says that this usually happened in three ways: as the outcome of a political or economic crises, as a consequence of a high-profile critical report on the system's performance, or "the energy and input of a new political and strategic leader". Sustaining improvement is associated with sustainment in teaching and pedagogy. "New pedagogy" is sustained on the basis of a set of beliefs among teachers about professional attitudes, values and dispositions. The authors note that they found that there were three ways

in which improving systems commonly do this: by establishing collaborative practices between teachers within and across schools, by developing a mediating layer between the schools and the centre, and by architecting tomorrow's leadership (ibid., 21). "Many systems in our sample have created a pedagogy in which teachers and school leaders work together to embed routines that nurture instructional and leadership excellence. They embed routines of instructional and leadership excellence in the teaching community, making classroom practice public, and develop teachers into coaches of their peers. These practices are supported by an infrastructure of professional career paths that not only enable teachers to chart their individual development course but also make them responsible for sharing their pedagogical skills throughout the system. In general, collaborative practices shift the drive for change away from the center to the front lines of schools, helping to make system improvement self-*sustaining*".

Interestingly the "new pedagogy" is not described substantively, but as the outcome of a process of collaborative practice and "peer learning" at school level and by a supporting context at above school level. This above school supporting context comprises "targeted on-hands support to schools", a buffer between the school and the centre, and a channel to share developments across schools. "As our sample systems have moved through their improvement journey, a number have chosen either to delegate responsibility away from the center to a newly created mediating layer located between the central educational authority and the schools themselves (e.g. school clusters or subject based groups), or have expanded the rights and responsibilities of an existing mediating layer (e.g. school districts/regions)." (p. 22)

The notion of performance stages and matching improvement strategies, also touched upon in BCG policy recommendations for Victoria and addressed in the leadership study by Day et al. (2009), is highly developed in this report. From the perspective of the educational effectiveness knowledge base there appears to be a more convincing rationale for the improvement potential of the first two performance stages (from poor to fair, and from good to fair) than for those mentioned for the higher stages. A contradiction might exist between the clear role for the centre in orchestrating the reform, and the supposed devolvement of authority to intermediary levels, when systems are operating at higher performance levels. Finally, the empirical support for the effectiveness of teacher cooperation and professional learning communities is quite modest (e.g. Lomos et al. 2011; Thurlings and Den Bok 2014).

The Scope of School Improvement in Comparison to Educational Effectiveness Research

Scholarly work on system and school improvement goes beyond application of the knowledge base on educational effectiveness. It goes beyond in the sense of taking a more holistic outlook, incorporating normative frameworks and persuasive presentation, and application of "high inference" conjectures on the basis of qualitative research and case study descriptions. In spite of this broader outlook it can still be

questioned on the empirical support for key mechanisms, concerning effective teaching, creating opportunities to learn, effective application of evaluation and feedback methods and efficient cooperation. Prescriptions for effective systemic reform deserve to be scrutinized on the compatibility of purposeful alignment orchestrated from the centre and the propagation of autonomy and "ownership" at school and teacher level. In some of the more visionary and future-oriented models high strung expectations of organization and governance arrangements take precedence over more "down to earth" levers for improvement, such as initial teacher training and curriculum alignment. In some cases overstressing secondary school organizational conditions, like creating the above school buffers and support structures, developing network structures with other schools, and schools expected to do research could be considered as putting at risk the primary tasks of schools, i.e. fostering student learning. The enthusiasm for the higher stages of development in the McKinsey taxonomy should not make us lose sight of the fact that the broadly supported practices of school internal accountability, as the key mechanism for the initial development stages, meets with many difficulties in practical application, like unfamiliarity of teachers to interpret test results, lack of didactic training to design action plans, resistance to use tests and time-consuming new administration and registration tasks. The dysfunctional facets of rational techniques and implications of alternative theories on the way schools function (Chap. 11) paint a more sobering reality of schools in a "coping" rather than a "continuous improvement" mode.

Conclusions

When making up the balance about the empirical results on effective schooling, one could say that despite sometimes disappointing effect sizes of quite sensible factors, like school-based evaluation and assessment, there is strong consensus on what could be described as a "positive theory of good schooling", featuring sufficient subject-matter exposure, achievement orientation, a positive school climate and a type of leadership that facilitates these characteristics.

One of the strong affirmations of the positive theory of good schooling is the fact that failing schools distinguish themselves on the same factors as effective schools: where successful schools do well on the identified effectiveness enhancing variables, failing schools do badly on the same factors. To quote Sam Stringfield: "A positive outlier school would have both a social and an academic climate of trust and sharing and talking positively about kids and solving their problems together A negative outlier school might or might not have a strong social cohesion among teachers, but typically had pretty strong norms that in your classroom you were the queen/king. Whatever worked (or didn't) for you wasn't relevant to the rest of the school, and you minded your own business on the work front". … "There are knowable steps toward the creation of a less effective school, and in the Louisiana study, we got to watch them play out over 11 years. The first step is to hire a

principal who lacks drive and commitment to constantly building and re-building an academically focused school. The second is to accept whatever teachers are sent or recommended to you through the system's HR office. On the long term, things go down- hill from there. We published an article on the differences in 1st year teacher's behaviors, fall to spring, in positive and negative outlier schools. Depressing, but clear. New teachers in more effective schools get better over their first year. In less effective schools, they don't. So the culture replicates itself, if un-attended" (Sam Stringfield, personal communication; Kirby et al. 1992).

A systemic perspective on educational effectiveness and school improvement, defined in a multi-layered structure of "given" and malleable conditions, has to consider the "steering potential" at each level. Research results tend to suggest that, in terms of variance explained and effect sizes, teaching processes at classroom level have somewhat more impact on student performance than school organizational conditions like leadership approaches, which in their turn are somewhat stronger than system-level levers of educational improvement. As stated by Wang et al. (1993) proximal conditions at micro level have stronger impact than distal conditions at macro level. To make a fuller comparison between the potential for improvement at the three levels (system, school and classroom teaching) two other, fairly obvious considerations need to be taken into account. Potentially, system-level policies will have implications for a much larger range of units (schools and teachers and ultimately students) than policy measures and professional activity at school and classroom level. Next, steering measures at higher levels are much "heavier" in terms of changing them, than are measures with a school-level or classroom-level scope. When, for example, an educational system wants to change from a tracked system for secondary education to a comprehensive one, this would take a heavy legislative procedure, most likely also political compromise and considerable implementation costs. Clearly this is a bigger and more encompassing issue than a school director who wants to introduce a pupil monitoring system, in all grades of a primary school.

Systemic reform is not just about identifying system-level levers for improvement, but also very much about orchestrating improvement-oriented measures at school and classroom level. According to the model that was introduced in Chap. 1, this "orchestration" of effectiveness enhancing conditions at lower level becomes more complex to the degree that there is more autonomy at school and classroom level. The way system-level policy measures are expected to bear effect then becomes a matter of "mutual adaption" and "co-construction" rather than straightforward implementation. The question of effective systemic reform becomes a question of alignment versus loose coupling.

In the theoretical parts of this book, variations of the rational planning model were seen as alternative steering models, differing in, let us say, their tolerance for autonomy at lower managerial levels. Proactive synoptic planning was seen as putting most of the demands on top-down policies and expectations of straightforward implementation. Exploiting market mechanisms was depicted as the strategy, most tolerant of strong autonomy at lower managerial levels. "Retroactive planning", exploiting the cybernetic principle of evaluation and feedback, has a

middle position; it is centralistic on output control and leaves freedom at lower managerial levels to choose processes and apply inputs.

Systemic reform in education is built upon assumptions of alignment between policy, management and professional teaching activities. According to certain normative principles, such as subsidiarity (doing nothing at higher levels that can be accomplished at lower levels), or according to given contextual conditions, such as historically grown patterns of centralization and decentralization, deliberate choices can be made for what might be indicated as "weak alignment" in the functioning of managerial levels (system, school and classrooms). An example of weak alignment is a general central curriculum "frame", consisting of broad objectives and general targets.

The review of the empirical evidence in educational effectiveness in other parts of this book suggest a "lean" model of school improvement. Key conclusions of this review established strong equifinality at teaching level, i.e. many different approaches all roughly equally effective, leaving room for summary constructs like "focussed teaching" (Louis et al. 2010) and "active teaching" (Hattie 2009). At school level a result-oriented approach, including formative assessment, feedback and stimulating test preparation and opportunity to learn were seen as the backbone of a "positive model" of school effectiveness, with instructional leadership, task-related cooperation between teachers and exploiting composition effects of good teachers as organizationally supportive conditions.

Substantively credible and simple scenarios for systemic reform are: (a) the *teacher training and recruitment scenario*: selective recruitment of teachers, teacher training providing sufficient support for teacher knowledge, particularly pedagogical content knowledge, all supporting effective teaching, with classroom management, support and cognitive activation, as the most proximal stimulant of improved student performance; (b) *the accountability scenario*: setting standards (as in a national core curriculum), developing and applying well-aligned summative tests and evaluations, possibly supported by more minute, well-aligned monitoring systems for formative evaluations, and opportunity to learn and test preparation (content aligned to standards and examinations) as the most proximal condition to stimulate student achievement; (c) *the good governance scenario*: patterns of functional decentralization fitting to the historical and cultural context of the country in question, lean, professional governance and school management provisions, including a basic "meta control" orientation, leaving room for distributive and "teacher leadership", facilitating task-oriented teacher cooperation and professional development; (d) *the evidence-based comprehensive reform scenario*: national stimulation of the implementation of evidence-based, comprehensive school reform programmes, in which school-level aspects of the first three scenarios are integrated.

Finally, despite good intentions all-over, a kind of worst case-scenario in school improvement could be discerned. This "scenario" is to be conceived as a set of loosely connected dysfunctional and inefficient elements. It might be labelled as the "*the displaced goal scenario*", as divergence from the primary goal of educational performance improvement is common in each of the elements. The main elements

are "policy churn", immunization against external monitoring and evaluation, organizational rationality of school support and trade-offs between improvement and modernization. *Policy churn* refers to frequent changes in government policies, which might imply that policy plans get insufficient time to be properly implemented or that new policy initiatives clash with ongoing policies in related domains, or just that there is an "overload" of new educational innovation activities (Keating 2015). In the case of "policy churn" the substantive rationality of performance improvement takes a back seat to the political rationality of politicians, who give priority to scoring in the policy arena and showing off with new ideas. *Immunization* is not to be excluded as one of the motives of schools in opposing external evaluation and monitoring. When schools use their autonomy to refuse external evaluations of government supported improvement programmes the whole improvement endeavour becomes obsolete. Apart from complete refusal, negotiations may lead to complicated compromise solutions; the case study on quality reform policies in the Netherlands, described in the next chapter, provides some examples. *Organizational rationality in school support* takes into consideration that support agencies may have vested interests in certain improvement strategies and remedies, or may be tempted to go for the "low hanging fruits" instead of the more fundamental levers for improvement. When autonomous schools manifest immunization tendencies with respect to sensitive areas, such as teacher appraisal or pupil monitoring, compromises may be reached to make external support "harmless" to schools and profitable for support agencies. Efforts that put secondary organizational processes (peer learning of teachers, "teachers as researchers", developing professional learning communities) at the centre deserve to be critically analysed from this perspective, where the key issue is the direct connection with improvement of the primary process of teaching and learning. Speaking of *trade-offs between improvement and modernization* is one way of drawing attention to the fact that improving educational performance in the sense of student learning outcomes is a clear and specific goal, while improvement-oriented action may also be driven by other interests. Goal displacement in this sense would mean that a *means,* in this case modernization of instruction or organizational administration, should not implicitly become a goal in itself. If modernization is an explicit goal, next to performance improvement, it could perhaps be defined as a *constraint* of performance improvement. Insufficient clarity about the improvement goal is, obviously, a core element of the "displaced goal scenario".

References

Ball, S. J., Mac Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. New York: Routledge.

Barber, M., Whelan, F., & Clark. M. (2011). Capturing the leadership premium: How the world's top school systems are building leadership capacity for the future. McKinsey & Company.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 133–180.

Bennis, W. G., Benne, K. D., & Chin, R. (1969). *The Planning of Change*. London: Holt, Rinehart & Winston.

Bishop, J. H. (1998). The effect of curriculum-based external exit exams on student achievement. *Journal of Economic Education, 29*, 171–182.

Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. H. (2014). International perspectives on teacher knowledge, beliefs and opportunities to learn. Teachers Education and Development Study in Mathematics, TEDS-M. Dordrecht: Springer.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research., 73*, 125–230.

Boston Consulting Group (2003). School Work Force Development Strategy Paper for the Department of Education and Training in Victoria. http://www.google.nl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCYQFjAA&url=http%3A%2F%2Fwww.education.vic.gov.au%2FDocuments%2Fschool%2Fprincipals%2Fprofdev%2Fpcpdschworkforce.doc&ei=K_MkVbLrC4XAPM_FgcAJ&usg=AFQjCNHmcEsxqnEWEDPBbGMHIhKJHU_DVg

Brophy, J. (1984). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631–661.

Brophy, J. (2001). Generic aspects of effective teaching. In M.C. Wang & H.J. Walberg (Eds.), *Tomorrow's teachers*. McCutchan Publishing Company.

Brunello, G., & Checchi, D. (2006). Does school tracking affect equality of opportunity? IZA Working Paper, Bonn.

Carnoy, M., Elmore, R., & Siskin, L. S. (2003). *The new accountability. High schools and high stakes testing*. New York: Routledge.

Causa, O., & Chapuis, C. (2009). *Equity in student achievement across OECD countries: An investigation of the role of policies*. OECD Economics Department Working Papers, No. 708. OECD Publishing. doi:10.1787/223056645650

Coffield, F. (2012). Why the McKinsey reports will not improve school systems. *Journal of Education Policy, 27*(1), 131–149.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2012). *Improving quality in education. Dynamic approaches to school improvement*. London: Routledge.

Cuban, L. (1990). Reforming again, again, again. *Educational Researcher, 19*, 3–13.

Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., et al. (2009). *The impact of school leadership on pupil outcomes*. Nottingham, UK: The National College for School Leadership.

Edmonds, R. R. (1979). *A discussion of the literature and issues related to effective schooling*. Cambridge: Center for Urban Studies. Harvard Graduate School of Education.

Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington D.C.: The Albert Shanker Institute.

Griffith, J. (2003). Schools as organizational models: Implications for examining school effectiveness. *The elementary school journal*, *102*, 349–366.

Hanushek, E. A., & Raymond, M. E. (2006). School accountability and student performance. Fedral Reserve Bank of St. Louis. *Regional Development, 2*, 51–61.

Hargreaves, A. & Braun, H. (2013). Data-driven improvement and accountability. Boulder, CO: National Education Policy Center. Retrieved (date) from http://nepc.colorado.edu/publication/data-driven-improvement-accountability/

Hargreaves, A., & Harris, A. (2005). Performance beyond expectations. National College for School Leadership. http://andyhargreaves.weebly.com/uploads/5/2/9/2/5292616/performance-beyond-expectations-summary.pdf

Hargreaves, A., Liebermann, A., Fullan, M., & Hopkins, D. (1998). *International handbook of educational change*. Dordrecht: Kluwer.

Harris, A., Chapman, C., Muijs, D., & Reynolds, D. (2013). Getting lost in translation? An analysis of the international engagement of practitioners and policy makers with the educational effectiveness research base. *School Leadership & Management*, *33*(1), 3–19.

Hattie, J. (2009). *Visible learning*. Abingdon: Routledge.

Hattie, J., & Alderman, E. M. (Eds.) (2012). *International guide to student achievement.* London: Routledge.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Hendriks, M.A. (2014). *The influence on school size, leadership, evaluation, and time on student outcomes*. Enschede: University of Twente, doctoral thesis.

Hopkins, D. (2002). Educational innovation: Generic lessons learned from (a) regional practice. Keynote Paper Presentation at the DECIDE International Seminar, University of Twente, the Netherlands.

Hopkins, D., Stringfield, S., Harries, A., Stoll, L., & Mackay, T. (2014). School and system improvement. *School Effectiveness and School Improvement, 25*, 257–281.

Keating, I. (2015). UK-England VO-Council, the Netherlands, European Federation of Education Employers (EFEE) and European Trade Union Committee for Education (ETUCE). Professional Autonomy, accountability and efficient leadership. Utrecht: VO Council.

Kirby, P., Stringfield, S., Teddlie, C., & Wimpelberg, R. (1992). Teacher induction in more and less effective schools. *School Effectiveness and School Improvement, 3*(3), 187–203.

Klieme, E. (2012). Qualities and effects of teaching. Towards a conceptual theory of teaching. Key note address. EARLI SIG meeting, Educational Effectiveness, Zuerich, August 23.

Kohn, A. (2004). Abusing research: The study of homework and other examples. *Phi Delta Kappan, 88*(1), 8–22.

Koretz, D. (2005). Alignment, high stakes, and the inflation of test scores. *Yearbook of the National Society for the Study of Education, 104*, 99–118.

Kyriakides, L., & Creemers, B. P. M. (2012). School policy on teaching and school learning environment: Direct and indirect effects upon student outcome measures. *Educational Research and Evaluation: An International Journal on Theory and Practice, 18*(5), 403–424.

Leithwood, K., Jantzi, D., & Mascall (1999). Large-scale reform. What works? Ontario Institute for Studies in Education, December 1999.

Levin, H. M. (2006). *Why is educational entrepreneurship so difficult?*. Columbia: Teachers College, Columbia University.

Locke, E. A., & Latham, G. (2002). Building a practically useful theory of goal setting and task motivation. *The American Psychologist, 57*(9), 705–717.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). The relationship between departments as professional communities and student achievement in secondary schools. *Teaching and Teacher Education*, *27*(4), 722–731 10 p.

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement*, *21*(3), 315–336.

Luyten, H., Scheerens, J., Visscher, A., Maslowski., R., Witziers, B., & Steen, R. (2005). *School factors related to quality and equity. Results from PISA 2000.* Paris: OECD.

Marzano, R.J. (2003). *What works in schools. Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as Myth and Ceremony. *The American Journal of Sociology*, *83*(2).

Miles, M. (1998). Finding keys to school change: A 40-year Odyssey. In A. Hargreaves, A. Liebermann, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change.* Dordrecht: Kluwer.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs: Prentice Hall.

Miron, G., Mathis, W. J., & Tornquist, E. (2010). Schools without diversity: Education management organizations, Charter Schools and the Demographic Stratification of the

American School System University of Colorado at Boulder: Dept. of Educational Leadership, Research & Technology College of Education.
Mitchell, C., & Sackney, L. (2000). *Profound improvement: Building capacity for a learning community*. Lisse: Swets & Zeitlinger.
Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset: Open Books.
Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better.* McKinsey and Company.
Muijs, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H., & Earl, L. (2014). Teaching effectiveness. A state of the art review. *School Effectiveness and School Improvement, 25*, 231–257.
Murphy, J. (1993). Restructuring schooling: The equity infrastructure. *School Effectiveness and School Improvement, 4*(2), 111–130.
North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.
OECD (2010a). *PISA 2009 Results Volume I, What students know and can do: Student performance in reading, mathematics and science*. Paris: OECD Publishing.
OECD. (2010b). *Strong performers and successful reformers in education: Lessons from PISA for the United States*. Paris: OECD.
OECD (2014). *PISA 2012 Results, Volume IV, What makes schools successful? Resources, policies and practices.* Paris: OECD Publishing.
Orbach, E. (1998). *How to assess the capacity of client government to implement an education development project*. Washington: World Bank.
Pannecoucke, I. (2005). *Gewikt en gewogen: Schoolkeuze tussen principe en feiten.* (*School choice between principles and facts*). Antwerpen: Universiteit van Antwerpen.
Patton, M. Q. (1978). *Utilization focused evalution*. Beverly-Hills: Sage.
Quinn, R. E., & Rohrbauch, J. (1983). Spatial model of effectiveness criteria, towards a competing values approach to organizational analysis. *Management Science, 29*, 363–377.
Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., Teddlie, C., et al. (2014). Educational effectiveness research (EER): A state of the Art Review. *School Effectiveness and School Improvement*, *25*, 197–230.
Sahlberg, P. (2011). *Finish lessons*. New York: Teachers College Press.
Sarason, S. (1990). *The predictable failure of educational reform: Can we change course before its too late*. San Francisco: Jossey-Bass.
Scheerens, J. (2012a). Mixed methods: An example of a school leadership effects study. Invited speech, University of Roma 3, Rome, May 21, 2012.
Scheerens, J. (Ed.) (2012b). *School leadership effects revisited. Review and meta-analysis of empirical studies.* Dordrecht: Springer.
Scheerens, J. (2013a). What is effective schooling. A review of current thought and practice. Washington, D.C.: International Baccalaureate Organization. Retrieved from http://www.ibo.org/globalassets/publications/ib-research/whatiseffectiveschoolingfinal-1.pdf
Scheerens, J. (2013b). The ripples and waves of educational effectiveness research: some comments to 'Getting lost in translation'. *School Leadership & Management, 33*(1), 20–25.
Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier Science Ltd.
Scheerens, J., Glas, C., & Thomas, S. M. (2003). *Educational evaluation, assessment and monitoring. A systemic approach.* Lisse: Swets & Zeitlinger.
Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organisation and Management.
Scheerens, J. (Ed.). (2014). *Productive Time in Education. A review of the effectiveness of teaching time at school, homework and extended time outside school hours*. Dordtrecht: Springer.

Scheerens, J., Luyten, H., Glas, C. A., Jehangir, K., & Van den Bergh, M. (2014). System level indicators. Analyses based on PISA 2009 data. Internal report. Enschede: University of Twente.

Scheerens, J., Luyten, H., Van den Berg, S. M., & Glas, C. A. W. (2015). Exploration of direct and indirect associations of system-level policy-amenable variables with reading literacy performance. *Educational Research and Evaluation, 21*, 15–39.

Schein, E. H. (1985). *Organizational culture and leadership: A dynamic view*. San Francisco: Jossey Bass.

Schmidt, W. H., Cogan, L. S., & McKnight, C. C. (2010). Equality of opportunity: Myth or reality in U.S. schooling. *American Educator, Winter 2010–2011*, 12–19.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499.

Slavin, R. E. (1996). *Success for all*. Lisse: Swets & Zeitlinger.

Slavin, R. E. (1998). Sands, bricks, and seeds: School change strategies and readiness for reform. In A. Hargreaves, A. Liebermann, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change*. Dordrecht: Kluwer.

Stringfield, S. C., Reynolds, D., & Schaffer, E. C. (2011). Toward high reliability public schooling. In R. H. Bellamy (Ed.), McREL: High reliability organizations in education. Denver: McNell.

Teddlie, C., & Stringfield, S. (1984). The Louisiana School effectiveness study. AERA-paper, New Orleans.

Ten Bruggencate, G., Luyten, H., Scheerens, J., & Sleegers, P. (2013). Modeling the influence of school leaders on student achievement: How can school leaders make a difference? *Educational Administration Quarterly, 48*(8), 699–732.

Thrupp, M. (2008). Secondary teaching, social contexts and the lingering politics of blame. Keynote address to the New Zealand Post Primary Teachers Association professional conference, 'Secondary Teaching on the Move', Waipuna Hotel and Conference Centre, Auckland, April 17, 2008.

Thurlings, M., & Den Brok, P. (2014). Teachers work as equals. What works? Eindhoven: Eindhoven School of Education.

Visscher, A., & Ehren, M. (2011). De eenvoud en complexiteit van opbrengstgericht werken. (The power and complexity of result oriented working.) The Hague: Ministry of Education. www.rijksoverheid.nl/bestanden/documenten-en-publicaties/rapporten/2011/07/13

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*, 249–294.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*, 1–19.

Weiss, C. H. (1980). Knowledge creep and decision accretion. Knowledge, Creation, Diffusion, Utilization. 1, no. 3.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39* (3), 398–425.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2009). *School accountability, autonomy and choice around the world*. Cheltenham, UK: Edward Elgar.

Chapter 13
Case Study: Quality-Oriented Educational Policy in the Netherlands

Keywords Bottom up reform · Evidence based educational policy · System evaluation · School evaluation · Teacher appraisal · Student assessment · Corporatism · School self-evaluation · School inspection · Accountability · High stakes testing · Equity oriented policy · Resilient students · Pupil monitoring systems · Social dialogue · Quality agendas · The principle of "Freedom of Education" in the Dutch constitution

Introduction

In this chapter description of structural arrangements and quality-oriented educational policy in the Netherlands will be used to illustrate educational reform and improvement "in practice". The chapter starts out with a brief sketch of the school system in the Netherlands, and some important structural characteristics, like the fact that two-thirds of the schools are government-dependent private schools. Other features are the highly differentiated nature of the school structure at secondary level, the existence of a strong middle layer of governance and an important educational support structure. In the subsequent sections two types of structural arrangements for quality-oriented policy are described. The first is the "social dialogue" among stakeholder organizations at the intermediate level, and the second is the structure for educational evaluation and assessment in the Netherlands. Next two policy programs are described. The first program, "Schools have the Initiative", is expected to bring the nationally agreed quality agendas into practice by an approach that is "owned" by the autonomous schools. The second program is the educational priority program, aimed at enhancing the position of disadvantaged groups of learners. The way the structure operates and the two programs are getting

This chapter is based on five reports by the current author, Scheerens et al. (2011, 2012), Scheerens (2013, 2014). These reports were contracted by the Dutch Science Foundation (NWO), the Inspectorate of Education/Ministry of Education, the VO Council and the European Union.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8_13

shape is compared to overall performance indicators, based on international studies. The results of the case study description are compared to some of the improvement-oriented policy scenarios described in the previous chapter, particularly the "good governance" and the "accountability scenario". The key question that is answered is whether the actual quality level of Dutch education, and prospects for its maintenance or further growth, is attributable to autonomy or to quality-oriented policy that is orchestrated from the centre.

Description of the School System

The Structure of the Dutch School System

Education is compulsory from the age of 5 to the age of 16; but pupils can (voluntary) enter primary education at the age of 4.

The Dutch education system is divided into three levels: primary, secondary and tertiary education. These levels include the following types of education:

- Primary education:
  - primary education for children aged from 4 to 12 years and
  - special primary education for children (aged 3–12) who require special educational treatment; for older pupils in this category, there is also special secondary education.

Primary schools in the Netherlands cater for children from 4 to 12 years of age. They are usually arranged into eight year groups.

Children in need of special care and attention can attend special schools.

- Secondary education:
  - secondary education caters for pupils between 12 and 18 years.

Secondary education is divided into as follows:

- pre-vocational education (VMBO) and individualized pre-vocational education (IVBO), 12–16 years;
- junior general secondary education (MAVO), 12–16 years (the MAVO was abolished in 1999 and integrated into pre-vocational education);
- senior general secondary education (HAVO), 12–17 years and
- pre-university education (VWO), 12–18 years.

Evidently, the system of secondary education is strongly stratified. After primary education, pupils move to one of the types of schools described above. Selection is informed by teachers' advice and achievement levels within primary education. The brightest students attend HAVO and VWO, and the rest goes to VMBO schools. From 1993 onwards, a shared curriculum during the first three years of secondary education students was the official policy (the so-called "*Basisvorming*" or basic

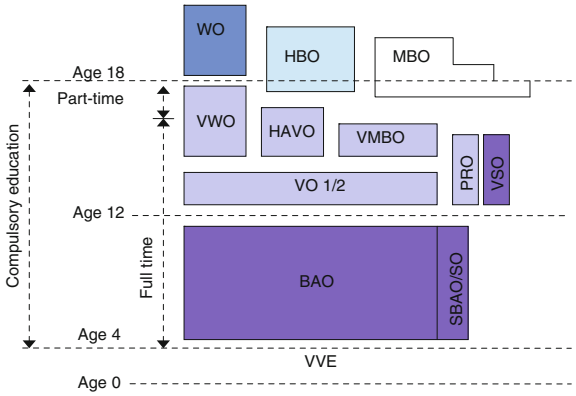

Fig. 13.1 Structure of the Dutch school system. *BAO* Mainstream primary education, *HAVO* General secondary education, *HBO* Higher professional education, *MBO* Vocational education, *PRO* Practical training, *SO* Special education, *VMBO* Pre-vocational secondary education, *VVE* Early childhood education, *VWO* Pre-university education, *WO* University education

general education). The introduction of basic general education could be seen as an attempt to introduce comprehensive schooling. Despite this attempt, the reality in most schools is streaming in which the differentiation at the upper secondary level is already "foreshadowed". In this respect, the attempt to introduce comprehensive schooling was not successful. The failure of the "*Basisvorming*" is documented in the report of the Parliamentary Inquiry Committee Educational Innovations, in 2008.

At upper secondary level, the differentiation consists of a HAVO, a VWO and a vocational stream of upper secondary education (see below). Because, in this book, the focus is on primary and general (academic) secondary education, further details on upper secondary vocational schools, tertiary education and adult education will be omitted, although their structure is shown in the diagram of the system, provided in Fig. 13.1.

High Stakes Testing and Examination at Primary and Secondary Level

At the end of the primary school program, most primary schools (in 2012 this amounted to 85 % of all primary schools) administer the "school leavers' test primary education" (*eindtoets basisonderwijs),* developed by CITO (the Central Institute of Test Development). The prime function of the "CITO test", as it is usually called, is to support the choice of track at the secondary level. The general acceptance of the CITO test should be understood from the way it functions with regards to the strongly stratified secondary school system. Research has shown that

the test score closely matches the actual advice that students get by the school. Driessen (2011) reports a correlation of 0.86 between test score and the advice by the school. The test score closely matches the actual advice that students get by the school. Studies from De Boer et al. (2007) and Driessen and Mulder (1999) show no indications of migrant students in the Netherlands receiving a lower or higher advice for a track in secondary education, compared to native Dutch students.

Gradually, the CITO test, i.e. the school aggregate score, is also being used as a basis for school evaluation. School scores on the CITO test are used for school inspection, and as a basis for school evaluation by municipalities and school boards.

At the end of each of the three tracks of secondary education, students sit an examination. The examination covers all subjects that the students have followed; students choose different subject matter packages. The examination consists of a school examination and a central examination. Schools are free to define the form of the school examination. Sometimes it is an aggregate of earlier accomplished tasks and partial tests, and sometimes it is a specific testing event, designed by the school itself, or bought from an external examination institute. The central examination takes place at a fixed day, and is externally administered. An arithmetic test is planned to be added to the central examination, but the precise function is currently, November 2015, still being debated.

A current trend is to weigh the central examination heavier than the school examination in determining the passing or failure of a student. Students have to reach a minimum standard in the basic subjects on the central examination in order to pass.

Division of Responsibilities

The Dutch school system is characterized by a large segment of government-dependent private schools. About two-thirds of all primary and secondary schools are government-dependent private schools, often based on a certain religious denomination. School governance and responsibilities for specific domains of functioning are largely similar for the public and (government-dependent) private schools, and differences do not appear to be relevant for the subject of this report.

As it comes to the division of responsibilities between administrative levels, schools in the Netherlands are among the most autonomous in the world. In 1998, 73 % of all decisions in areas like instruction, planning and structures, resources and human resources were taken by the school, 4 % by the local level and 24 % by the central level (source Education at a Glance, OECD 1996, p. 299). In 2008, the percentage of decisions taken by the school had risen to 94 %, whereas only 6 % of the decisions were taken at the central level (source: Education at a Glance, OECD 2008, p. 488). School autonomy in the Netherlands is very high in all functional domains that were considered in the OECD study: organization of instruction, planning and structures, personnel management and resources. It is important to

note that setting examinations is one of the areas in which the central level remains ultimately responsible, although the technical process of developing examinations is delegated to organizations like CvE (Committee for Examinations), CITO (Central Test Institute) and SLO (Foundation for Curriculum Development).

School autonomy and "freedom of education" is also manifested in the domain of curriculum development. At the central level the so-called "core objectives" are established. For example, for secondary education 58 core objectives divided over all subject matter areas have been formulated. These core objectives leave a lot of freedom for schools to apply them in their own school curriculum. In the recent past the "Education Council" (*Onderwijsraad*), which is the central advisory committee, has made a case for more specific standards. A compromise solution in this direction is the introduction of so-called "reference levels" for language and arithmetic. This development might be seen as an emerging trend towards more detailed specification of educational objectives; although this is a very sensitive issue in the Netherlands (e.g. Education Council 2012).

Financing of general secondary schools was adapted in 1998, when very detailed and school specific forms of school finance were replaced by block grants. The lump sum per school is determined on various ratios, all of them ultimately depending on the number of students. Applying the budget has to abide by certain conditions of financing, which are subject to accountancy control.

Some Specific Features and Relevant Trends

School autonomy in the Netherlands can be seen as rooted in the principle of "freedom of education", as a part of the constitution, since 1917. The practical implication of this constitutional law was to create equality between state schools and private denominational schools. The structural component of this policy was an unofficial intermediary power structure, made up of denominationally grounded organizations and pressure groups that represented teachers and school governors. Some analysts (e.g. Leune 2007) have described this structure as corporatist. Part of this intermediary structure is a substantial educational support structure, partly organized according to denomination as well. This support structure has a particular position in the whole national governance structure for education. As one would see it as a "techno structure", it should be realized that it is an instrument in the hands of educational organizations and individual schools rather than an instrument of the central government, i.e. the Ministry of Education. During the last decade, the intermediary structure of educational organizations has been concentrated and secularized in the form of the PO, VO and MBO Councils. These councils have actually taken on the role of employers in their respective sectors. Additionally, they also support schools in matters of improvement and innovation.

Major trends in educational governance in the Netherlands during the last 20 years are a shift towards more school autonomy, a retreat of central government, particularly in areas like finance and personnel policies and a transformation of the

intermediary level of educational organizations, with still a lot of influence residing at this level. In areas like curriculum, accountability and quality control, there has been a slight tendency to take back some of the autonomy to the central level. Particularly, in areas of defining and evaluating final outcomes of schools, central government has recently implemented additional legislation on performance standards/reference levels and centralized national testing. This latter trend is manifested in the quality agendas of the Ministry of Education, more prescription and weight of central examinations, and more specific "end terms" in the sense of reference levels in basic subjects. The Inspectorate of Education has been a more constant factor in these slowly changing organizational arrangements, holding its central role of supervising the quality of education, notwithstanding the fact that in the Netherlands the quality of education is seen as a responsibility of the individual school, i.e. the school board. Bronneman (2011) gives a more detailed analysis of the changes in educational policy making structures over the last 20 years (1990–2010). She notes, among others, that the influence of the central government, i.e. the Ministry of Education, has diminished in important ways; that devolving authority to school boards has not created more autonomy for teachers and that teachers have lost authority in the new governance order.

The Network Structure of Quality-Oriented Educational Policy in the Netherlands

Autonomy is often associated with the enhanced quality of education. In the previous chapters, we have seen that the empirical knowledge base does not really support this position. Although there are some positive results, the overall picture is quite dubious. Some theoretical approaches (e.g. Chubb and Moe 1990) have postulated intermediary mechanisms that might explain the connection between school autonomy and educational outcomes, like more responsiveness to the local community, less administrative burden and unleashing creativity in the organization. We have seen that the Dutch educational system is characterized by high school autonomy, a generally devolved educational structure, an important "mid field" of educational organizations and stakeholders, but also by a tradition of high stakes examinations, and more centralized accountability concerns. In this section, more details are provided on how this "constellation" functions, how processes of quality-oriented policy evolve and which positions are taken by the various stakeholders.

When it comes to governance and quality-oriented policy in the Netherlands, the involvement of institutional stakeholders can be described in terms of a social dialogue between employers of education, the central government and trade unions (EFEE 2012). In this section, quality agendas and the way these are getting shape through this social dialogue are described.

How the Social Dialogue Functions

The social dialogue in education is officially organized in the Foundation for Education (*Stichting van het Onderwijs*), since 2010. This platform has been founded by the social partners. The following organizations take part: the PO Council (PO-Raad), VO Council (VO-raad), Council for Secondary Vocational Education (MBO-Raad), the Association for Higher Education (uniting tertiary-level institutions at ISCED 5B level), called VSNU, and trade unions in the field of education, united under two main umbrella organizations, FNC and CNV. Next, there is a number of *associated organizations*, representing, among others students, and the *Educational Cooperation*, which is a joint organization of the trade unions in education.[1]

The issues addressed in the Foundation for Education are influencing national educational policy, addressing educational themes that transcend the various education sectors and active contribution to a positive image of education. Although the Foundation for Education is the formally institutionalized body where the social dialogue can take place, in actual practice important agreements are often established on the basis of collaboration between partners beyond the Foundation. The broad scope of issues that are addressed is evident from National Agreements on Education (*Nationaal Onderwijsakkoord*), between the members of the Foundation and the Minister of Education. The version of the Agreement which was signed in October 2013 covers an extra spending on education, salary levels and conditions of labour for teachers, training and professional development requirements, "professional space" for teachers, as well as extra financial means for the stimulation of educational quality. In addition to agreements on sector over-arching themes, covered by the Foundation for Education, at sector level (primary, secondary, vocational and higher education), more specific agreements are made between employers in education and trade unions on the one hand and the Ministry of Education on the other. For example, the Policy Agreement for Secondary education (*Bestuursakkoord VO*) has "Improvement of education quality" as its main theme. This agreement was reached in September 2013 and it is oriented towards better student performance, improved quality and professional development of teachers and school leaders (source: http://www.vo-raad.nl/dossiers/bestuursakkoord-vo). A similar agreement has been reached between the PO Council and the Ministry of Education (source: http://www.edux.nl/site_files/uploads/BESTUURSAKKOORD%20PRIMAIR%20ONDERWIJS%202012.pdf). This agreement has a strong orientation to improvement of educational quality as well, and, like the *Bestuursakkoord VO*, contains explicit performance indicators to be reached by 2015.

[1]AOC raad, Interstedelijk Studenten Overleg (ISO), LAKS, Landelijke Studenten Vakbond (LSVb), Nuffic, Onderwijscooperatie, Platform Vakinhoudelijke Verenigingen Voortgezet Onderwijs, Vereniging Toezichthouders Onderwijsinstellingen (VTOI) Source: http://www.stichtingvanhetonderwijs.nl/over_de_stichting/geass_organisaties.shtml.

Quality Agendas Issued by the Central Government and Accorded by the Social Partners

The Dutch central government has explicit quality agendas for primary, secondary and upper secondary vocational education. These "quality agendas" are to be seen as based on the Act on Primary Education (WPO and WEC), and the Act on Secondary Education (WVO). Quality agendas exist for primary, secondary general and secondary vocational education, formulated by the Ministry of Education in (2007a, b, 2008).

Improved quality in basic school outcomes (language and arithmetic/mathematics) is a common feature in the three quality agendas. Next, new types of achievement outcomes are intended, citizenship skills and competencies in secondary education and work related competencies in vocational education. Attainment targets, in the sense of decreased early school leaving, are also included in the quality agendas. The proper functioning of teaching processes and schools as organizations are "throughput" issues that are emphasized as well. These are monitored on the basis of the work of the Educational Inspectorate, which rates schools as regular, weak or very weak. The quality agendas also mention certain general classes of levers for quality improvement; these are testing, achievement monitoring, examinations and quality care measures, teacher training, continuous professional development, evidence-based innovation and better alignment among different school types and education levels.

It should be noted that these acts and government-initiated programs are to be seen as specifying general targets and conditions for quality enhancement, while the ultimate responsibility for educational quality resides with the schools. Requirements on "Good Education and Good Governance" which are part of the Law on primary and secondary education set additional conditions on the governance structure of primary and secondary schools, which are expected to contribute to high educational quality, including student achievement outcomes.

More operational programs that are to be seen as vehicles to reach the targets of the quality agendas, and action plans for better performance, are the programs "Schools have the Initiative" (*School aan Zet*), the "Teachers' Agenda" (*Lerarenagenda*) and various initiatives to support "evidence based" educational policy.

A Closer Look at the Quality Outlook of the Social Partners

Judging from the Agreements as described in the above, there is a general agreement between social partners and the Ministry of Education on the priorities of the quality agendas. This incorporates acceptance of educational outcomes as key performance indicators, as well as agreement on a number of conditions (good governance, strengthening of the teaching profession) that are to be seen as supportive.

Despite of this overall consensus, one could say that there is a certain caution among some of the trade unions with respect to the above-school establishment of general learning standards (reference levels), and particularly to the summative use of performance assessment. For example, in an (undated) note of the Algemene Onderwijs Bond (AOB) the position is taken that deregulation and increased school autonomy have given more power to school managers but constrained the professional autonomy of teachers (source: http://www.aob.nl/kixtart/modules/absolutenm/articlefiles/6048-Profstat.pdf).

From the teaching field, a lot of criticism is ventilated against educational testing which, according to these critics, has gone over the edge, and is exaggerated. At primary school level, an action group "Save primary education" expressed rabbinate opposition against mandatory performance testing.

Another line of debate is taken in a recent report by the Education Council, which is the most important general advisory committee to the government, in which the current quality policies are described as "narrow" (Onderwijsraad 2013: "A narrow view on Education Quality"). The council sees negative side effects of the current concentration on basic subjects and educational testing of these basics (e.g. teaching to the test). Instead, broader "Bildung", non-cognitive competencies and skills as well as "twenty first century skills" are emphasized by the council.

In the Policy Agreement for Secondary Education[2] (already cited in the above), the VO council underlines the role of school leaders in enhancing the quality of education: "Good teachers are the prime agents in quality enhancement. Good school leaders enable teachers to do a good job, when they align HRM policies, finance and school organization with pedagogical and didactic goals" (p. 5). The VO Council also states that it is currently active in the development of instruments to improve the professional development of school leaders (p. 8), such as a mandatory training for school leaders, instruments for self-assessment and personal development. Specific targets for school leader competencies and competency development are formulated in the Agreement (p. 19). A competency profile for school leaders is described in the report "*Competency Profile for School Leaders in Secondary education*" by Andersen and Krueger, commissioned by the VO Council. The competency model that is presented has five core competencies, and confirms to the category of integrated school leadership models.

The concept of "professional space" addresses the relationship of teachers and school leaders; the latter are expected to safeguard the professional autonomy of the former. On this particular issue, there appears to be agreement with the educational cooperation, in which the teacher unions are represented.

In the Policy Agreement for Primary Education (2012–2015), the PO Council mentions the quality action plans and teacher policies (Teachers 2020) as joint ambitions between the Ministry of Education and the Council. The aim of the Policy Agreement is enhancing student performance in primary education. Part of

[2]http://www.vo-raad.nl/userfiles/bestanden/Bestuursakkoord/20111214-Bestuursakkoord-VO-raad-OCW.pdf.

the Agreement entails the delegation of authority to school boards, putting the PO Council in the position of educational employer; this is described under the heading of "More space for school boards". Like the Agreement for Secondary Education, the professional development of school leaders, including competency requirements is one of the explicit targets. By 2015, a school leader register should be operational. In 2013, about half of the school leaders in primary education had registered (source: http://www.poraad.nl/files/bestuursakkoordgetekend.pdf).

The PO Council has founded a "School leaders Chamber" (http://www.poraad.nl/content/schoolleiderskamer). In this Chamber school leaders from 30 school boards, which are members of the PO Council, meet to discuss topics of common interest, examples are, among others "result oriented leadership". The PO Council currently organizes master classes on governance and leadership.

Trade unions for teachers and other educational organizations in the Netherlands cooperate in the education cooperation. This body, raised in 2011, deals with the central government on issues that regard the position of teachers.

In May 2012, the education cooperation signed a Policy Agreement with the Secretary of State for Education. For the government, the education cooperation is seen as the partner and council for all issues that involve the professional quality of teachers. The teacher register, in which teachers can register their competencies, is one of the prime achievements of this Policy Agreement.

On the issue of "professional space" teacher policies meet school leader policies, as trade unions tend to be weary of limitations of teacher autonomy due to stronger involvement of school managers and school governors. ETUCE, as the international organization of trade unions in education, has formulated an explicit vision on school leadership. In this vision, the professional autonomy of teachers is guaranteed on the basis of leadership styles that respect and exploit the professional autonomy of teachers (ETUCE 2012).

The Role of School Boards

In the Netherlands school boards are the "competent authorities" of the schools.

School boards are constituted locally and there is little information nationally regarding the qualifications of school board staff; school governors may be volunteers or professionals. They may be parents of students in the school, citizens from the local community, members of a religious or life philosophy community or professionals with specific expertise such as law, finance, HRM or education. Hence, especially on smaller school boards, there may not be any staff with an educational background. The recent (2010) legislation on "Good Education, Good Governance" has significantly strengthened the role of school boards for quality enhancement (Scheerens et al. 2011). This has also led to the strengthening of the role of the national "intermediary" or umbrella organizations, the primary and

secondary education councils (PO-raad[3] and VO-raad[4]), which are the national associations of schools boards of primary and secondary education.

In a review of educational evaluation and assessment in the Netherlands, (Nusche et al. 2014), the OECD examiners write the following about the functioning of schools boards:

> "School boards particularly have very uneven capacities and this is exacerbated by the fact that these bodies may have very different legal statuses, sizes and scopes of responsibility given the fact that those who have created and are supervising them (e.g. religious communities or municipalities) enjoy large freedom in the way they regulate their remits and operational procedures. The members of school boards are typically lay persons who may lack the capacity to conduct evaluation with appropriate professionalism and the size of their permanent staff is often quite small. While some school boards formulate ambitious performance demands and exert pressure on their schools to improve, others seem to be less committed to quality goals and do not make appropriate efforts to raise quality ambitions".

In one of the few Dutch empirical studies on the functioning of school boards, Hooge and Honing (2014) state that the "Good Education, Good Quality act " holds school boards ultimately responsible for guaranteeing educational quality. They state that almost all primary schools are now publicly financed and privately run (from local government to school boards), 46 % run just one school, 34 % run two schools and 20 % run more than ten schools.

They estimate that school boards consist for about 50% of parents, mostly highly educated, mostly men. Two-sixths of the school boards consist of no parents, but of at least one professional governor, supplemented with voluntary governors. School board members are appointed, and therefore are more like trustees rather than representatives. Appointments occur by cooptation. Professional governors (paid) are appointed by their internal supervisors (*Raad van Toezicht*). According to Hooge and Honing, schools in the Netherlands lack democratic accountability. In their study, they focus on a basic prerequisite of school boards being capable of managing educational quality, by raising the question whether school boards have a picture of the quality of the service delivery of their school. They conclude that "in general" school boards (in primary education) are aware of the quality of their school.

Evaluative Comments

The way the educational social dialogue has been institutionalized in the Netherlands is a clear example of the Dutch "*Poldermodel*", in which organized stakeholders and interested parties "from the field" come to terms with the national government on all kinds of issues. The strength of this model is agreement and

[3]See the website of PO-Raad (The Primary Education Council) here: http://www.poraad.nl.

[4]See the website of VO-Raad (The Secondary Education Council) here: http://www.vo-raad.nl.

commitment from practically all interested parties. A weakness, according to some critics, is the time-consuming nature of the decision-making process.

Although there is good communication and broad agreement on main aspects of educational quality, there are definitely also disagreements and conflicts, particularly when it comes to assessment and monitoring of quality targets.

Trade unions act against what they see as a diminishing of the professional autonomy of teachers, because of increased external control, a diminishing of trust and "being reduced to a measuring instrument". From the position of school leaders/school governors, there is fear for negative side effects of accountability policies, in the sense of schools becoming more selective in student admission, in order to look good on the outcome indicators.

Although the structure seems to be made to negotiate and reach compromises, there are also inherent tensions in two arenas: power and influence on the one hand, and antagonist positions on empirical analytic approaches to quality enhancement, versus more interactionist visions on bottom-up reform. In quality-oriented educational policy these tensions become visible, for example when some analysts still think that the role of the centre should be further diminished, while others conclude that bottom-up reforms have frequently failed (concrete examples of the latter are given by Bronneman 2011, 447). We shall turn back to this issue in subsequent sections, when describing the role of evaluation and assessment in more detail, and when discussing two more specific programs, "Schools have the initiative", and equity-oriented policy.

The Functioning of Educational Evaluation and Assessment in the Netherlands

Context and History of the Evaluation System

As stated in the previous sections, a key feature of the educational system in the Netherlands is the principle of freedom of education, which should be seen as the historic background for the particular patterns of central and de-central elements in educational governance and the strong autonomy of schools.

During the last three decades, school autonomy has also grown in areas like financial management (the introduction of block grants and lump sum financing) and personnel policy, and a continued effort is being made to deregulate and decrease central administrative pressure. As it was quoted before, currently, according to publications from the OECD, the Netherlands is one of the countries with the highest degree of school autonomy in the world.

When it comes to positioning and describing evaluation and assessment, these historically developed structural and institutional arrangements are of great relevance. Types of evaluation range from being embedded in centralistic control measures to internal "formative" assessment procedures that are seen as instrumental to the development of individual students and schools.

The history of evaluation and assessment in the Netherlands goes back to the 1970s, when there was a temporary upsurge in more government-driven educational innovation policy, the so-called "constructive educational policy", led by the social democratic Minister of Education, Van Kemenade. Next to having a somewhat more centralistic orientation, a second major characteristic of this policy was a rational planning approach. Large-scale innovations were planned as experiments, and scientific evaluations were to point out the viability of the innovations. In many ways, this approach did not go well with the traditional autonomy of schools, and neither did it sit well with the intermediary structures, including the denominationally organized educational support organization which had to manage the innovatory projects. The efforts to scientifically evaluate these programs largely failed due to unclear organizational positioning of the evaluation researchers and particularly due to strong resistance from teachers and schools.

During the 1980s, educational policy changed to a more incremental system-wide development orientation and was matched by an evaluation approach that left the model of program evaluation, turning to a more "systemic" evaluation approach, based on key data streams and continuous monitoring. In this period major instruments for system-level evaluation were developed: the periodic assessment project (PPON), the cohort studies in primary and secondary education as well as a gradual development of policy-relevant educational statistics and educational indicators, the latter strongly stimulated by the active participation of the Netherlands in the OECD education indicators project (INES) and in initiatives from the EU, more particularly EURYDICE.

In the wake of these developments in system-level evaluation, systematic student evaluation and school evaluations (in that order) were gradually getting off the ground. Despite of the large autonomy of schools, the Netherlands traditionally has had central examinations at the end of secondary schools. In primary schools, a school leavers test, the so-called CITO test, is being used as a basis for supporting the choice of a secondary school track, since 1976. During the 1980s, school inspection was taking a turn towards more empirical analytic approaches, like quality indicators and structured classroom observations. Partly based on knowledge from school effectiveness research, and also partly following similar orientations of Her Majesties' Inspectorate in the United Kingdom, an elaborate supervision framework was developed, and applied in school inspection.

Currently, educational evaluation and assessment is being aligned to evolved models of school governance, with slightly changed roles for national government, intermediary organizations and autonomous schools, giving rise to new forms such as "risk based inspection", "windows for accountability" (an instrument for internal school inspection) and "formative assessment in function of result oriented work". A second important policy development has come forward in the form of the, already mentioned, quality agendas and Action Plans of the Minister of Education. In these policy plans, clear attainment targets and benchmarks are set for the improvement of educational achievement, particularly in basic school subjects. A specific orientation for schools is propagated under the title "Result oriented work". These new policy emphases underline the importance of summative and

formative student assessment, and instrumental feedback to improve teaching and learning. Making the primary school leavers test obligatory also fits within these policy plans.

The Functioning of Evaluation and Assessment

All instances of evaluation, assessment and appraisal addressed confirm to two basic elements of the definition of systematic evaluation: they involve structured empirical data and allow for an evaluative judgment. Scheerens et al. (2012) distinguish three major evaluation functions: certification, accountability and improvement/organizational learning. When crossing these dimensions (functions) with two other dimensions, namely type of data and aggregation level, a taxonomy of evaluation types can be drawn up, featuring a total of 14 evaluation types. Practically, all of these types are applied in the Netherlands, which means that the scope and application of evaluation and monitoring is quite broad. Among this broad range of evaluation approaches there is one blank spot, which means that it is a form that may occur, but on which little systematic information is available, and that is teacher appraisal. Teacher appraisal in the Netherlands belongs to the discretion of the competent authorities of schools, i.e. school boards and municipalities. Just a few monitors exist on the degree to which and the way how schools fulfil this evaluative function. The fact that this domain is not penetrated by external organizations, not even the Inspectorate, can be seen as one of the purest features (or toughest strongholds, when one takes a more critical perspective) of professional autonomy within autonomous schools.

Expertise in evaluation technology is well developed in the Netherlands. The test development company CITO has an international reputation in advanced applications of educational testing. Next, a range of research institutes and university departments exists that have the research technical skills to carry out various forms of educational evaluation. In the past, the development of evaluation has been stimulated by national expert committees, such as the Committee for Program Evaluation, and the research school (network of universities) on educational research, ICO. Last but not the least the systematic approaches of the Dutch Inspectorate have often been cited as exemplary by other educational Inspectorates in Europe.

In the next section, the functioning of system, school, teachers and student evaluation will be briefly summarized.

Overview of System, School, Teacher, and Student Evaluation

The landscape of educational evaluation and assessment in the Netherlands is schematically described in two overviews included in the Annex to this chapter: the

first one providing a short description of the type of evaluation, formal responsibility and information on implementation and use of the evaluation and assessment results. The second overview adds the main instruments used for each evaluation type and the periodicity of its administration. In both schematic overviews, a distinction is made between system evaluation, school assessment, teacher appraisal and student assessment.

More detailed descriptive information about these evaluation types is provided in Scheerens et al. (2011), Scheerens (2013) and OECD, Nusche et al. (2014).

System Evaluation

After the developments concerning the evaluation of national innovatory programs and the gradual move to a monitoring type of systemic evaluation, as described in the above, a number of stable data collection procedures were established. These are the PPON, the cohort studies, the gradual development of national educational statistics to a set of "system indicators" (Dutch: *bestel indicatoren*) and the various data collections by the Inspectorate. In addition, the Netherlands has been taking part in international assessment studies, like TIMSS and PISA, since the 1960s. Finally, a large number of smaller scale evaluation studies are being contracted out by the Ministry of Education. It should be noted that system-level evaluation partly depends on information that is primarily collected for purposes of pupil assessment or school evaluation. Examples are examination results, aggregate data based on the Cito primary school leavers' test and aggregate information from the inspection of individual schools.

Several reporting frameworks have been created, in which sub-sets of these data are synthesized, annually. These are the publications "Trends" and "Key Data", and the Inspectorate's annual report.

School Evaluation

The main instruments for school evaluation are school self-evaluation and school inspection. School self-evaluation has a legal basis in the obligation of schools to produce various annual documents, among which school plans. In addition, school self-evaluation can be seen as part of quality care, which is one of the quality aspects that are monitored by the Inspectorate. In the early 2000s, school self-evaluation got a strong boost by the installation of two dedicated bodies (called Q5 and Q Primair), which stimulated school self-evaluation practices in primary and secondary schools. In addition, school self-evaluation was stimulated by the school inspection strategy, indicated as "proportional inspection". In the Educational Supervision Law of 2002, proportional inspection was described as a basis for determining the degree of intensity with which schools were to be inspected. If schools had quality care and self-evaluation in good order, they were to be inspected less frequently.

From the late 1990s onwards, school inspection became more systematic and guided by explicit supervision frameworks, in which quality aspects and quality indicators were defined. The Inspectorate also issued quality cards, in which school functioning was rated on a number of indicators. School quality cards were made publicly available and were meant to support school choice.

At about 2005 new concepts on educational and school governance brought about changes with respect to both school self-evaluation and school inspection. The legislation about "Good Education, Good Governance" (Parliamentary year 2004–2005) set about to create new governance relationships, intended to give more space and responsibility to schools, create further deregulation and to diminish administrative burdens. The school boards' responsibility for educational quality was underlined, urging for a clear delineation of supervision tasks, which ultimately led to the prescription for the creation of separate supervision boards. In correspondence with these changes, the role of internal supervision and horizontal accountability by schools was underlined and distinguished from external supervision and vertical accountability, through school inspection. The adapted governance arrangements opened the possibility for a new interpretation of selective inspection, which played into the stronger positioning of school boards and their responsibility for horizontal accountability. Risk-based inspection was initiated, from 2006 onwards, and consists of a first screening of all schools, based on a relatively limited set of information sources (among which educational achievement indicators), on the basis of which one of two inspection arrangements is chosen. An arrangement can be basic (no risks for the quality of education), or adapted (weak or very weak quality). Next, more intensive supervision and improvement stimulation occurs for the schools that are classified as (very) weak. Apart from receiving support, weak schools are also urged to improve by the threat of sanctions, which may go as far as holding back the complete budgetary funding of the school.

School inspection and school evaluation in general are expected to benefit from value-added performance measurements, which can be based on existing and prospective instruments for student evaluation and are currently explored in pilot projects.

Internal supervision and horizontal accountability are currently supported by a new procedure in which the VO and (very recently) the PO Councils cooperate with the central data unit of the Ministry of Education (DUO) in "Windows for Accountability". Through this procedure, schools obtain core statistical information on their own functioning through DUO and are supported to create school-based indicators on parent satisfaction with the school. This development might be seen as a more structured and externally supported stimulation of school self-evaluation, as compared to the more autonomous arrangements of earlier periods, which had somewhat disappointing results.

Teacher Appraisal

In the Netherlands, the evaluation or appraisal of individual teachers belongs to the jurisdiction of the Competent Authority of the school, the school board or the municipality.

Although the central role of teachers is fully recognized in current educational policy, there is no external teacher appraisal. Effective teaching is an important issue in school inspection, but it does not regard the functioning of individual teachers.

Few evaluative studies are available in which the way school boards and school leaders appraise teachers is described and evaluated.

Student Assessment

Centrally specifying educational objectives and testing them by means of national tests is a theme that is not uncontroversial, given the principle of freedom of education and the traditional strong autonomy of schools in the Netherlands. Despite of the sensitivity of this issue, the Netherlands has a central examination at the end of secondary education. For a long-time educational attainment, targets (Dutch: *eindtermen*) were only described in rather general terms. A fairly recent development is the formation of somewhat more specific "reference levels", or benchmarks for language and arithmetic. An important step is also the decision to implement a national school leavers test, implemented in 2014, and to make pupil monitoring systems in primary and secondary schools mandatory.

The three most important instruments for student assessment in the Netherlands are the CITO school leavers test at the end of the primary school period, the secondary school examinations, which consist of a central and school-based part, and pupil monitoring systems in primary and secondary schools, the most important of which are also developed and supported by CITO.

The quality agendas and action plans for better achievement stimulate the use of formative and summative student assessments. A very interesting development that is a consequence of these policy initiatives is the formative use of achievement tests, part of pupil monitoring systems, to stimulate result-oriented work. One could say that it is particularly at this micro-level of teaching and learning that the improvement potential of assessment is at stake in a very concrete way. Experiences so far are promising, but also point at strong needs for professional development and external support to teachers, in order to learn to work effectively with information from tests.

Responsibilities for Evaluation and Assessment

System-level evaluation is mostly controlled by the Ministry of Education. Cohort studies are a joint venture of the Ministry of Education, the Central Bureau for Statistics and the organization for scientific research (NWO).

In school evaluation, autonomous schools have an important say. The recent legislation on Good Education, Good Governance, has underlined the responsibilities of the school board for quality enhancement as such, and internal supervision in particular. School-level external supervision is the responsibility of the Inspectorate. The educational organizations, united in the VO and PO Councils, have a supportive role in stimulating internal school supervision and horizontal accountability.

Teacher appraisal fully belongs to the discretion of the competent authorities of schools, i.e. school boards and municipalities.

With the exception of the central examinations, student assessment belongs to the jurisdiction of schools. Instruments like the CITO school leavers test at primary level and the pupil monitoring systems at primary and secondary levels are purchased by schools. Although the application of these instruments has become (as in the case of monitoring systems) or is becoming mandatory (as is the case of the primary school leavers' test), schools still decide about the particular instrument they want to use.

Implementation, Appreciation and Use of Evaluation and Assessment

Implementation of evaluation and assessment procedures in the Netherlands has sometimes hampered because of lack of cooperation from schools in data collection procedures. This was noted with regards to the early program evaluations in the 1970s, and also in the difficulties for the Netherlands in obtaining sufficient response rates in international studies. For this reason, the Netherlands was excluded from the international reporting on PISA 2000 and the first wave of the TALIS study (about the functioning of teachers). School autonomy and a general weariness of administrative burden might explain this phenomenon. Still, autonomous schools have bought into important student assessment instruments like the CITO school leavers test and pupil monitoring systems, on a very large scale.

With regards to the implementation of school self-evaluation, a mixed picture emerges. It is the impression that schools generally have acquired school self-evaluation instruments, including administrative systems. Yet, the proportion of schools which, according to the Inspectorate, have their quality care in order is rising slower than was expected.

Systematic information on the appreciation of evaluation procedures is only available with respect to school inspection. Generally, schools are satisfied with the work of the Inspectorate. An internal review by the Inspectorate pointed out that the recent risk-based inspection is successfully being implemented and has shown results in the sense of a diminishing number of very weak schools.

The notion of evaluation and assessment stimulating the improvement of teaching and learning works differently for evaluation procedures at system, school, teacher and student levels. In a general sense, all types of evaluation and

assessment, both summative and formative, are ultimately meant to improve educational achievement through improved teaching and learning. Feedback loops and improvement mechanisms will differ, however, both in length and the role of different actors in using evaluative information for improvement purposes.

With regards to the use of system-level evaluations, there is only fragmented and anecdotal evidence available. The availability of periodic synthetic publications such as the annual report of the Inspectorate and the publications on Trends and Key Figures must be seen as an important condition for facilitating use. Since the reporting of the Parliamentary Committee "Dijsselbloem" in 2008, public interest in the position of the Netherlands on international assessment tests, such as PISA, seems to have grown and has been the object of some debate in the press.

Research studies point out that use of school self-evaluation procedures by schools is often superficial and that schools experience problems in actually using information from self-evaluation for improvement purposes.

Similarly, recent and ongoing studies into schools implementing the ideas of "result oriented work" are pointing out that effective use of data from student achievement tests requires skills and experiences that are often lacking among teachers. The good news is that these practices can be considered as touching the core of what evaluation and assessment can do for improving teaching and learning, and that current improvement and evaluation policies in the Netherlands are addressing this very core.

Evaluative Comments

With respect to system-level evaluation, the Netherlands has a broad range of procedures and instruments available and the continuation of these seems to be guaranteed.

At school-level evaluation, procedures are being aligned to new governance arrangements, which may lead to more efficient arrangements, a smaller burden and more effective support to schools.

Teacher appraisal is weakly developed. As far as student assessment is concerned, the most important policy initiative is probably the stimulation of result-oriented work by schools. Among others, this is a strong stimulant of the formative use of results from achievement testing and pupil monitoring by teachers and schools.

Nusche et al. (2014) give the following overall assessment of educational evaluation and assessment in the Netherlands:

> "In many ways, the evaluation and assessment policies developed in the Dutch school system are in line with the principles identified by the OECD to develop a comprehensive and balanced evaluation and assessment framework. The Dutch evaluation and assessment approach stands out internationally as striking a good balance between school-based and central elements, quantitative and qualitative approaches, improvement and accountability functions and vertical and horizontal responsibilities of schools".

One of the suggestions the OECD reviewers offer is to "*Embed the evaluation and assessment framework with broader learning goals, hinting at the development of a system of long-term learning goals. These learning goals are to be placed at the center of the framework and all other evaluation and assessment activities are to be aligned with these goals*" (ibid., 13). Not surprisingly, another recommendation is to further develop teacher appraisal. The recommendation to design a more explicit framework of learning goals invites to see the functioning of evaluation and assessment as embedded in the larger policy and organizational context of the Dutch system. The recommendation makes a lot of sense from a rational analytic perspective, but in the actual Dutch context there is quite a history that documents why such a central framework of learning goals does not exist. The principle of "Freedom of education" makes explicit learning goals a sensitive issue, which is readily associated with "state pedagogy", a term that is almost a curse for some of the political parties. The position of the central advisory body to the Minister of Education the *Education Council* on this issue and how the Council's position has shifted over time is also quite telling. In the late 1990s, and the first years of 2000 the Education Council made a strong case for explicit learning standards. As documented by Bronneman (2011), these proposals were effectively blocked in the period that the Christian Democrats party held the Ministry of education. Only after a sharp increase in the number of complaints about the lack of basic skills in arithmetic and language, a step in the direction of more concrete learning goals was taken, by implementing the so-called reference levels in language and arithmetic in 2010 (ibid., 426). However, even these reference levels are fairly broad, and tests to measure them, particularly the arithmetic test for secondary education, have not yet been introduced. To complete the picture the Education Council, as the first champion for learning standards, seems to have shifted its position since 1999 considerably. In a recent advice with the title "A narrow view of quality" (*Onderwijsraad* 2013), the council sees cognitive testing as endangering other more pedagogical aims in schooling.

Apart from the established core of central examinations, high stakes testing and mandatory formative assessment, evaluation provisions are part of the struggles for power and influence, in which the central government seems on its retreat and the organized middle field and other stakeholders are gaining influence (Bronneman 2011). This has led to a double set of evaluation provisions, some of them external and government controlled, and others owned by the school boards (Scheerens et al. 2012; Scheerens 2013).

Evidence-Based and Bottom-up Reform in the Netherlands

Tensions Between Autonomy and Quality

Tensions between autonomy and quality-oriented educational policy manifest themselves most strongly with respect to mandatory testing and the implications of the ambitions with respect to "evidence based" reform. Since the end of the 1990s,

autonomy was further increased, and the interplay between the government, educational organizations (particularly those representing school boards) and autonomous schools was summarized in a steering philosophy for which the term "governance" was applied. According to this philosophy, "the government tries to organize the relationships between students, parents, teachers, school boards, municipalities and the Inspectorate in such a way that they correct one another in keeping a balance and stimulate one another to innovation and entrepreneurship" (Frissen et al. 2013, p. 4). The idea of "governance" is that schools not only make themselves accountable to the government, but especially to parents and the public at large (Ibid., p. 4). In actual practice, school autonomy is conditioned by the way the government seeks to stimulate the quality of education. This policy does not only specify outcomes but also stimulates certain approaches that are seen as instrumental to realizing improved outcomes, such as mandatory testing and the approach of "result oriented work". Frissen et al. (2013) conclude that the ideal of governance, in the sense of actors operating in a balanced structure through a system of checks and balances, is not realized, because the position of teachers and parents is much weaker than that of school boards, intermediary organizations and the government.

When taking the set of government initiatives to enhance the quality of education as the focus, one could characterize the playing field by observing that central quality norms and preferred work approaches (such as "result oriented work") might be seen as limiting school autonomy, while, at the same time, school autonomy could be seen as constraining the possibilities of the intended "evidence based" way of reform and school improvement.

This will be further illustrated by comparing two partly differing and partly overlapping approaches to realize the quality agendas and action plans for better performance. The first one is an agenda for effect evaluations of policy interventions, under the heading "Outlook on Effectiveness" (Dutch: *Zicht op Kwaliteit*), and the second is the "Governance Agreement", regarding support for the implementation of major facets of the overall quality enhancement policy (such as result-oriented work). The latter approach is bottom-up reform, which is covered in the program "Schools have the Initiative".

Outlook on Effectiveness

"Outlook on Effectiveness" lists a total of 50 current policy interventions that have been analysed for their being amenable to effect evaluation by the Central Planning Bureau.[5] Of these 50 interventions, 34 were considered evaluable. A more in depth

[5]Ministerie van Onderwijs en Wetenschappen (2012) Stand van Zaken Zicht op Kwaltiteit. Internal report.

study of a more limited number of policy interventions (Van Elk et al. undated)[6] indicates that in actual practice, there are many problems in realizing strong evaluation designs; for example, in quite a few cases the interventions had already started, so that a base line measure was not feasible. In a review of studies that assessed quality enhancement in primary schools, Scheerens and Doolaard (2013) concluded that agreed plans for randomized field trials could not be carried out, because of the lack of cooperation from schools. "Outlook of Effectiveness" enforces the idea of evidence-based policy, where policy plans are scrutinized for research-based support and evaluated with strong research designs. In actual practice, this approach seems to experience difficulties first because policy interventions are implemented before science-based analyses and baseline measures have had the chance to take place, and second because autonomous schools are hard to be persuaded to take part in experimental studies. Other programs and structures that were set up to realize evidence-based educational policy are the formation of a Top Institute on Evidence-Based Research (TIER) and an academic research program, likewise dedicated to practical topics for which strong evidence is sought (Prove for Education, Dutch *Onderwijsbewijs*). It is beyond the scope of this presentation to discuss these programs, also given the fact that intended evaluations have not been completed.

The Program: "Schools Have the Initiative"

"Schools have the Initiative" is based on an agreement between the Ministry of Education and the PO and VO Councils. In this agreement, the partners state that they give space and confidence to school boards to do the right things. Each school can make its own choice in improving its quality. Schools will not be burdened with additional accountability requirements. The responsibility lies with the school boards, which are expected to use their regular channels for external review, such as the annual school report.

"Schools have the Initiative" is a three-year program (2012–2015) for school boards and schools in primary and secondary education, aiming for higher educational outcomes within the framework of national ambitions. "Schools have the Initiative" have translated these national ambitions into program goals in six domains. According to the philosophy of "Schools have the Initiative", schools ought to be the "owners of their change processes". The six domains of "Schools have the Initiative" are as follows:

1. Achievement-oriented work;
2. HRM/learning organization;

[6]Van Elk, R., Van der Meer,F., Van der Stee, M. & Webbink, D. (undated). Zicht op Effectiviteit. Studie naar Evaluatie Ontwerpen voor Onderwijs en Wetenschshapsmaatregelen.

3. Basic skills (language and arithmetic in primary schools, and core subjects English language, Dutch Language and Mathematics in secondary schools);
4. Dealing with differences between pupils;
5. Excellence/gifted students and
6. Promotion of Beta skills, science and technology.

More specifically the program "Schools have the Initiative" intends

- to stimulate school boards and schools to formulate ambitions and goals;
- to provide school boards and schools with relevant knowledge and expertise, and give them insight into good practices, relevant parties and resources and how to get "from A to B"; and
- to keep a sharp eye on the progress that boards and schools make (given their own goals and ambitions), and provide transparency at national level.

The actual program is based on schools soliciting a budget from the Ministry to carry out work within one or several of the domains, and attracting external experts, from a national pool of experts and from the existing educational support infrastructure.

"Schools have the Initiative" aims for participation of 3000 primary schools, 450 regular secondary schools and 150 secondary schools offering special education. In November 2012, a total of 2800 schools (both primary and secondary schools) were involved.[7]

At this time the division of school projects over the six domains was indicated in Fig. 13.2 (same source).

Result-oriented work (OGW), dealing with differences between students and the Learning Organization, appeared to be the mostly chosen domains at the time. By early 2015, the emphases had shifted to a predominance of developments in the domain of the school as a learning organization.

According to Frissen et al. (2013), the annual budget for "Schools have the Initiative" is about 150 million per year for elementary schools, for a period of three years. For secondary schools about the same funding is available (bringing the total close to a billion EURO).

A core element in "Schools have the Initiative" is the way schools are supported by external experts. During the three-year period, each school that takes part is entitled to four meetings with an expert. The experts are indicated as either "ambition experts" or "thematic experts"; some experts are specialized in primary schools, and others in secondary schools and or special education http://www.schoolaanzet.nl/over-school-aan-zet/de-experts/.

When looking at the profiles of the "ambition experts" (primary schools), it appears that of the 29 experts the large majority has a background of organization and management consultant and 6 experts also have a background in didactics and teaching and learning. For the "thematic experts" (again, primary education) the

[7]Source: Informatory note for the Council of Directors at the Ministry of Education, November, 2012.

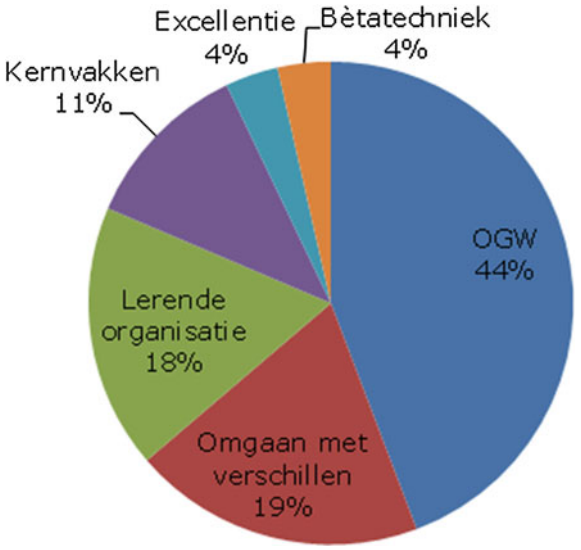

Fig. 13.2 Emphases in sub-domains of "Schools have the initiative" in November 2012 (OGW stands for result-oriented work, *Omgaan met verschillen* is dealing with differences between pupils, *Lerende organisatie* is Learning Organization, *Kernvakken* stands for developments in basic subjects, language and mathematics, *Excellentie* stands for excellence and attention for gifted students and *Beta-techniek* stands for emphases on science and technology

picture is reversed. Of the 43 experts who are listed, 33 have a background in subject matter didactics (language, arithmetic, mostly), and teaching and learning; and 10 others have more of a background of organization consultant, or HRD manager.

Facets of the program that are interesting from the perspective of effectiveness of systemic reform are the program evaluation that is planned, the role of experts as "critical friends" and the question where "schools have the initiative" stands in relation to the overall intention of evidence-based educational policy.

Program Evaluation

The Ministry of Education has contracted out a program evaluation of "Schools have the Initiative" that comprises process and effect evaluation.

The central research questions for process evaluation are as follows:

- Which kind of operations has been implemented in "Schools have the Initiative" and which instruments and means have been used for this?
- Is "Schools have the Initiative" being accepted by school boards and schools and how satisfied are the participating schools?
- How satisfied and committed are stakeholders belonging to the existing educational infrastructure?

The core questions for effect evaluation are the following:

- What is the starting position of schools that participate in "Schools have the Initiative" on core indicators (as stated in the agreement between the Ministry and the PO and VO Councils), as compared to non-participating schools?
- Can differential growth on the core indicators be detected, when comparing program and control schools?
- How did the implementation process in the program schools develop in the period between the pre- and post-measures, and how does the program implementation process influence the results of the previous core question, about differential growth between program and control schools.

As implied in the evaluation questions for the effect evaluation, a comparison group design was chosen for the evaluation. In the actual evaluation plan, a quasi-experimental design, where control schools are drawn from schools which enter the program one year later, was given preference over a randomized control group design (as this was considered as practically unfeasible).

The effect evaluation is to be based on quantitative indicators, related to participation objectives (number of participating schools) and realization of performance indicators that vary for the different sub-programs.

Examples of indicators are as follows:

- By 2015, all participating schools in primary and secondary education have identified their (top 20 %) excellent students, developed a targeted approach for these students, and visible performance effects of this approach (sub-program Excellence);
- In 2015, fixed percentages (ranging from 30 to 55 % for various types of secondary schools) choose a Beta or technical profile (sub-program science and technology);
- In 2015, all participating schools score significantly higher on (self-selected sub-domains) of basic subjects as measured by standardized achievement tests (sub-program achievement-oriented work);
- In 2015, all participating schools in secondary education have implemented a learning school organization, and are developing a targeted approach with respect to HRM and achievement-oriented work (sub-program Learning Organization and HRM); and
- In 2015, "practically all" teachers of the participating schools are sufficiently able to provide differentiation in their teaching and deal with individual differences between students (Professionalization in differentiation).

The actual program evaluation is carried out by a research institute. A striking characteristic is that the evaluation approach uses existing data streams to a large extent. Data sources like statistical data from DUO, process indicators from the Inspectorate of Education, achievement results from the COOL cohort studies and outcomes of various monitors, and even information from the register of teachers, are all intended to be used for measuring the quantitative indicators. Process evaluation partly uses more qualitative data and inventories of "good practices".

Critical Friends

Little documentation was found on the possible monitoring role of the experts involved in "Schools have the Initiative". The way the role of the experts is described is as informants, counsellors, supporters and advisors of schools, and not as evaluators, not even "friendly critics". In the way the experts present themselves on the web site, very few of them elaborate on specific monitoring and evaluation experiences and skills.

Evidence-Based Reform?

Initiated by the advice of the Parliamentary Committee on the Innovation of Secondary Education, the Committee "Dijsselbloem" in 2008, educational policy was to become evidence based. New educational policies would need to be piloted, and these pilots had to be rigorously evaluated, preferably by means of randomized field trials. "Schools have the Initiative" was launched at a time when scientific evaluation studies on result-oriented work were still in progress and results not yet available. Moreover, the program is much broader than result-oriented work, for which at least explicit rationales exist (Visscher and Ehren 2011; Scheerens and Doolaard 2013). Some authors express doubts about the programmatic activities of schools as part of "Schools have the Initiative" being evidence based (cf. Frissen et al. 2013).

Yet, "Schools have the Initiative" may be considered as a kind of hybrid form of evidence-based reform, combining expert input and central evaluation design with the traditional Dutch "bottom up" innovation approach. What is lacking in most areas (and some truly research-based approaches to result-oriented work are probably the only exceptions) are explicit program rationales and operational, integrated programs for reform, as, for example in the American Comprehensive School Reform programs. Perhaps, programs should be seen as being substituted by expert advice in "Schools have the Initiative", but this is likely to give rise to a rather fragmented implementation process, with difficulties for schools and clusters of schools to be supported by the "best evidence available". In terms of ex post evaluation, "Schools have the Initiative" is attempting a rigorous evaluation design, which is, however, vulnerable due to the great complexity of the program, with its several sub-programs. Ex ante evaluation, in terms of the specification of a program theory that incorporates state-of-the-art research-based evidence on educational effectiveness, is missing out. In the program description, the term program theory is used, but the description does not go beyond an overview of the various elements of the program.[8]

[8]Ministerie van OCW (2012) Nadere Offerte Aanvraag voor: proces- en effectevaluatie "School aan Zet".

Evaluative Comments

The combination of "Outlook on Effectiveness" and "Schools have the Initiative" can be seen as a typically Dutch construction of giving different stakeholders their share, an approach that was described earlier as the "Polder Model". A positive aspect of this combination of approaches is that indeed, the major stake holders (government, school boards, and the VO and PO Council) are all committed, with the educational support structure getting its usual share.

At the same time more critical questions might be raised:

First of all, as indicated in the previous paragraph, it is debatable whether the "Schools have the Initiative" approach meets the demands of the evidence-based approach.

Second, despite the rather positive assessments of evaluability by the Central Planning Bureau, some doubts are in place about of the feasibility of the evidence-based approach in the Dutch context. Ironically, school autonomy is sometimes getting in the way of evidence-based approaches because schools refuse to cooperate in data collection and controlled treatment implementation.

Third, critical questions might be raised about the efficiency of the improvement approach in "Schools have the Initiative". The Dutch tradition of bottom-up reform depends strongly on relatively small networks of schools who, often with consultants, support organizations, or universities, tackle complex issues, for example, how to deal with gifted children. In this way, even if good practices may be discovered, mechanisms for dissemination are mostly lacking. This is clearly a limit to the learning potential of the system, as far as improvement and innovation is concerned. Critical comments of the efficiency of this approach are also given in Bronneman (2011), Nusche et al. (2014) and Scheerens and Doolaard (2013).

Effectiveness of Equity-Oriented Educational Policies and Programs in the Netherlands

Main Equity-Related Policy Challenges

In the Netherlands, children from low SES (socio-economic status) families and first- and second-generation immigrants appear to be the most disadvantaged with respect to access, treatment and outcomes of education. This is evident from drop-out and attainment rates, a certain degree of inequality of resources and "treatments", and a broad range of outcome indicators (examples to be provided in the sequel).

The prime factor to be considered as possibly detrimental to equity is the strongly tracked secondary school system. The Netherlands scores extremely high on an "index of stratification", which is based on indicators like "age of first

selection for secondary schools", "number of school types with specific educational programs" and percentage of students in lower vocational education (Van Ravens 2011). The age of first selection to secondary education is 12 years old, while the number of secondary school tracks is no less than 7 (OECD 2012). The majority of students entering secondary education enrol in lower vocational schools (Dutch: vmbo schools). Apart from these structural facets of stratification, there are additional characteristics of Dutch secondary schools that represent selectivity: a high percentage of students is in schools where students' records of academic performance are considered for admittance, and an equally high percentage is in schools that group students by ability (OECD 2012, p. 57).

Inequalities are not associated with funding issues, since public and private schools in the Netherlands have equal funding. As will be documented in more detail below, special funding measures form the back bone of policies to overcome inequalities.

Apart from the main challenges regarding low SES students and students from immigrant background, there are no other major target groups that would require special attention of equity-oriented policy. Performance differences between boys and girls are moderate, as compared to other countries (Scheerens et al. 2013), and groups that were focused in early equity-oriented policies, namely children from inland shipping parents and travelling labourers, such as fare workers, have declined in importance over time. There is some attention for geographical areas where school performance is lagging behind, particularly in the North of the country (provinces Groningen and Friesland). However, these do not have much profile, given the strongly decentralized nature of equity-oriented policies in the Netherlands.

In order to illustrate these challenges, equity relevant indicators based on the PISA 2000 data base are presented in Table 13.1 (Luyten et al. 2005). In this report the nature and outcomes of equity-oriented educational policy in the Netherlands are considered over a period of roughly 40 years, starting from the 1970s until now. The information dated from 2000 is to be seen as a kind of mid-term assessment of the state of affairs. In the next section, more recent data will be presented on the degree to which educational equity-oriented policies are reaching the desired results.

Key indicators relevant for equity-oriented policies are

- indicators on the variability of educational performance;
- indicators on the degree to which student performance in cores subjects is influenced by the SES of the students and minority background;
- performance differences between students from migrant and non-migrant backgrounds and
- resource inequalities.

The indicators shown in Table 13.1 are cited from Luyten et al. (2005), based on PISA 2000 data, reading literacy achievement.

These figures illustrate that migrant status is an important challenge to equality of opportunity of students at lower secondary level in the Netherlands, according to these data from PISA 2000.

Table 13.1 Equity outcomes based on PISA 2000 data

Indicator	The Netherlands	OECD average
Between school variance as a percentage of total between student variance	52 %	36 %
Difference in score points between top and bottom quarter of the index of socio-economic status	71	92
Difference in score points between non-immigrant and immigrant students	83	58
Differences in score points between boys and girls	30	31
Percentage of teachers with an ISCED 5A qualification in the language of instruction in the bottom quarter of schools in SES composition	71 %	72 %

Main Policy Initiatives and Reform Measures

Priority Levels in Equity-Oriented Policy

The history of equity-oriented educational policy goes back to the 1970s. Initially, the sole focus was at mainstream primary education. Since the 1970s, pre- and "early school" (VVE) became additional priority levels. Secondary education has never been a focus level in equity-oriented policy, apart from special attention and some special programs dedicated to lowering drop out among students from cultural minorities in vocational education, MBO in particular. In a way, the vocational tracks of lower secondary education could be seen as a provision that caters to the needs of students from low SES status and cultural minorities. However, this statement is almost cynical as it points at the selectivity of the tracked school system. The official policy is that the barriers between the secondary school types are passable and that "upstream" to a higher school category is possible. In actual practice, this happens rarely and as a general issue the inflexibility and lack of horizontal mobility across the school categories is a notorious weakness of the Dutch secondary school system (cf. Van Ravens 2011).

Policy Initiatives and Reforms: "Treatments" to Enhance the Position of Disadvantaged Learners

Current and most recent developments in equity-oriented policies in Dutch primary and pre-primary education should be understood from a historical perspective spanning about 40 years.

Table 13.2 provides an overview of the main phases and emphases between 1974 and 2015.

The most constant policy instrument that has been used in the Dutch equity-oriented policy is the extra funding of schools based on school composition. Low SES and cultural minority students count as more than one student in the

Table 13.2 Phases in equity-oriented policies in Dutch education

Phase	Policy levers and instruments	Budget indications
Educational stimulation policy (1974–1982)	Special funding on the basis of pupil weights, educational support	100 million guilders per annum (source, Scheerens 1987)
Educational priority policy (1982–1998)	Same as above + educational priority areas (school networks and a regional support organization)	Extra for the area component: 22 million guilders (figure for 1986) (source, Scheerens 1987)
Educational policy oriented at disadvantages (1998 onwards)	Same as above + additional focus on the pre- and early school phase (VVE) Decentralization to municipality level	700 million EURO per annum. Source, Mulder and Meijnen (2013)
Additional components since 2010	Extra financing of so-called Impulse areas poverty zones Further decentralization to school board level Initiative to harmonize pre-school provisions Pull-out strategies (Dutch: "Schakelklassen" and "Startgroepen") Extended school time Differentiation component in general quality-oriented educational policy: "effectiveness by measure"	

formulas for the school budget, which are based on the number of students enrolled. The extra increment has varied over time from 0.03 to 0.08 and 1.2. The most recent "weights regulation" has two levels: 0.03 when parents have lower vocational education as their maximum educational level and 1.2 when one of the parents has just primary education and the other lower vocational education. Schools are eligible for extra funding when they have a certain percentage of students that are targeted by the educational priority of disadvantaged students policy; this threshold has been lowered from the 2006–2007 school year, namely from 9 to 6 %. Since 2010, eligible schools in the so-called impulse areas receive another increment in their budget over and above the student weight-based formula.

Schools are expected to spend the extra funding on measures that enhance the position of their disadvantaged learners, but they are free to decide how they do this (Implication of autonomy 1). Extra teaching and support staff, partly used for class size reduction, and buying in external support are the main "treatments" that schools are likely to choose. There is no precise information on this, because schools cannot be held accountable for how they spend their budget (Implication of autonomy 2). Finally, there are no evaluation studies that can attribute results of equity-oriented policies to the actual treatments that schools implement on the basis

of their extra funding, not only because of local control over the use of the funding, but also because of frequent refusals from school to participate in research and evaluation studies (Implication of autonomy 3). Thanks to a long tradition of cohort studies, outcomes that are most relevant to equity-oriented policies can be monitored quite well, but it is very difficult to find schools that are ready to cooperate in experimental or process–outcome evaluation studies (Scheerens and Doolaard 2013).

From 1998 onward, pre-school education (VVE) became a second major target area for equity-oriented policy, next to regular primary education. The policy levers are exactly the same as described above, extra funding based on the "weights regulation", and pre-school institutions and schools free to choose treatments.

Since 2010, extra measures have been stimulated by the government and key stakeholders, like employers of education. Different kind of pull-out strategies, where special classes are formed of eligible students who get extra treatment, like additional Dutch language education and extended learning time (longer school day, school week, or summer schools).

An approach by the Inspectorate of Education, known as "risk based" inspection, which is in the field since 2007, deserves specific attention. Within the framework of risk-based inspection, the Inspectorate uses annual risk analyses to target inspection visits to potentially failing schools. When a school is diagnosed as showing "weak" or "very weak" performance, intensified follow-up inspection and extra support for these schools follow. According to internal evaluations by the Inspectorate, this policy has been quite successful in diminishing the number of weak schools (De Wolf and Verkroost 2011).

Finally, equity stimulation is more recently getting an extra boost, as it is being profiled as a dedicated component of more general educational policy that is aimed at enhancing quality and performance "*Basis voor Beter Presteren*" (Mulder and Meynen 2013; Driessen 2013).

Assessment of the Effectiveness of the Policy Measures

Monitoring and Program Evaluation Studies

Throughout the history of equity-oriented policy, from the 1970s until now, attempts at program evaluation have been made. Invariable results of these program evaluations indicated no, or very small, improvement on outcome indicators regarding the performance of the two main at-risk groups, low SES students, and students belonging to minority groups (most of them from Surinam, the Dutch Antilles or from Turkish and Moroccan background). Moreover, when improvement could be noted, it was never possible to attribute these outcomes to implemented policy measures. For example, progress of minority students could largely be explained by the duration of being in the country, rather than to any educational factor (Scheerens 1987; Driessen and Mulder 1999; Van der Werf 1995; Mulder

and Van der Werf 1999; Meijnen and Mulder 2013; Driessen 2013). The main conclusions from the most recent evaluation study, the one by Mulder and Meijnen (2013), are the following:

– Pre-school education (VVE) has led to widely varying organization structures and substantive priorities. The authors conclude that decentralization and ever growing autonomy have caused an overall inefficient approach.
– The connection between pre- and early school (group 2 to group 3) is steadily improving although performance effects among students could not be established.
– The most recent modification of the "weights regulation" has not changed much, and has not led to a better position of rural schools (which was one of the aims). Including the educational priority means in the lump sum has blurred the picture on the spending of the dedicated budget. The authors fear that much of the moneys dedicated to enhancing equity are spent on irrelevant posts.
– Over the period between 1994 and 2007, minority students have caught up considerably in their general performance as compared to native Dutch students, but are still lagging behind. No progress could be noted in Dutch language performance, neither among minority students, nor among Dutch target group students (low SES) over the same period.

In the next section, as a second source of assessment, more elaborate use will be made of national and international equity indicators. Together, these indicators allow for a differentiated and nuanced picture of the state of affairs on educational equity in the Netherlands.

Results Based on National and International Indicators

Variability and SES determinacy of outcomes

- The Netherlands is rated as manifesting medium to high equity, when the total student performance variation is used as an indicator of dispersion, based on PISA 2009 data (Scheerens et al. 2013).
- The Dutch educational system is to be rated as low in equity when the between school variance is considered. PISA 2012 data show that the between school variance in the Netherlands amounts to 65 % of the total between student variation, while the OECD average is 40 %. For the within school variance, these figures are 35 % for the Netherlands and 60 % for the OECD average. This pattern of high between school variation and relatively low within school variation is directly associated with the tracked structure of secondary schooling (OECD 2014a, b, Vol II).
- The degree to which educational performance is influenced by the socio-economic status (SES) of the students is an important indicator of inequity. On this indicator the Netherlands does slightly better than the OECD average, results from PISA 2012 (OECD 2014b, Vol II p. 59). The percentage of

variance in mathematics performance explained by the ESCS index (educational social and cultural status) for the Netherlands is 11.5 %, as compared to the OECD average of 14.6 %.

- In subsequent waves of PISA, the percentages of the so-called *resilient* students have been calculated. Resilient students are those among the 20 % lowest on the national ESCS distribution that score in the 20 % highest scoring segment of the international PISA score distribution. The Netherlands is above average as far as the percentage of resilient students is concerned: 14.8 % in PISA 2006(science), 10.57 in PISA 2009 (reading) and 8 % in PISA 2012 (mathematics), with OECD averages, respectively, 13.1 (2006), 7.16 (2009) and 7 (2012); (Luyten 2011; Scheerens et al. 2013; OECD 2014b, Vol II).
- When analysing the performance of Dutch students at the low tail of the performance distribution in terms of absolute scores on the PISA performance tests, it appears that these students are doing very well, as compared to students in other countries. For example, the bottom 20 % of Dutch PISA 2006 reading literacy performance distribution belongs to the international top 5 countries. Judging from PIRLS 2006 results, the 25 % lowest scoring students on the national reading literacy performance distribution score in the international top 5 countries (Van der Steeg et al. 2011). At the same time, there is much concern in the Netherlands about the fact that Dutch students in the top 20 % segment do relatively less well; but this is seen as a concern about *excellence* rather than equity.

Minority students

- Luyten (2011), on the basis of Dutch longitudinal datasets, conclude that the educational achievement and attainment of the ethnic minorities in the Netherlands are showing slow but steady improvement. This is particularly true of the Turkish and Moroccan communities, although their overall educational level remains below that of the 'native' Dutch population. The improvement is large due to the fact that second-generation migrants have been able to complete their full education in the Netherlands.
- Between 2005 and 2008, the percentage of early school leavers showed a downward trend in all student groups, although ethnic students remained more likely to "drop out" than their Dutch counterparts. A situation that has proven particularly difficult to resolve is the high concentration of ethnic minority students in some schools in the large cities. During the 2006/2007 school year, ethnic minority students made up over 80 % of the student body in some 40 % of primary schools in Amsterdam and Rotterdam (Luyten 2011, p. 85).
- Results from PISA 2006 show that Dutch students from first- and second-generation minorities lag further behind than the average among OECD countries. For example, in mathematics second-generation minority students score, on average, 66 points on the PISA performance scale below the average of Dutch native students, as compared to an OECD average of second-generation minority students lagging behind 45 points (Luyten 2011).

- Results based on PISA 2012 show that the lagging behind of minority students as compared to Dutch native students amounts to 57 points, in comparison to an OECD average of 34 points. The disadvantage of minorities appears to have become slightly less over time though (with an improvement of 9 points between 2006 and 2012); (OECD 2014b, Vol II p. 227).
- Students from non-western immigrant groups are overrepresented in the lower tracks of upper secondary vocational education (MBO); Source: Ministry of Education (2012, p. 112). The participation rate in tertiary education of non-western immigrant students has slightly gone down between 2008 and 2012 (same source, p. 136). The participation of Dutch students with parents who have low education levels in higher education is 16 %, as compared to an OECD average of 17 % and an EU average of 18 % (Source: OECD 2012a, p. 102).

These indicators paint a diverse picture of the state of affairs with respect to equity in Dutch primary and secondary education. As compared to other countries, the Netherlands does not seem to do so badly on general indicators on dispersion (the total between students variation) and "SES determinacy" of student performance. When the Dutch results are compared with other systems that have tracked secondary school systems, like Belgium, Germany, Luxemburg and Austria, the Netherlands does relatively well, given the extreme diversification of the school system. Still the categorical system produces a high between school variance, and a strong influence of school composition, in the sense of the average SES in a school. Most positive with respect to equity is the high performance of students at the low end of the performance distribution.

The lagging behind of cultural minority students remains a serious issue, and improvement is quite slow.

When making up the balance, the following conclusion can be drawn:

Funding is the core "instrument" that has been used over a 40-year period in the history of educational policy to improve the position of disadvantaged learners in the Netherlands. It has appeared quite difficult to assess its effects in a precise way, because there is limited information about the way autonomous schools implement the extra budget.

Use of ICT is assumed to have played no role, or a very minor role, in the way schools have used extra funding. But basically, this is unknown, given the fact that schools are not held accountable on how they use extra funding. Moreover, ICT is not present as a "process indicator" among the quality indicators used by the Inspectorate.

As far as **key success system factors** of equity-oriented policy in the Netherlands are concerned, the main candidate is the consistent compensatory funding of schools catering for substantial percentages of disadvantaged learners, over a 40-year period. The overall effect of these funding arrangements could be hypothesized as the explanation of the relatively high achievement of Dutch students at the low end of the performance distribution as compared to other countries.

Systemic factors, influencing achievement in basic skills in the Netherlands, appear to be equity suppressing rather than enhancing. A tracked school system,

rather general curricular objectives, grade repetition, selectivity including admission patterns and inflexible pathways in the sense of horizontal and vertical transition between school categories reinforce rather than suppress underachievement associated with social background and minority status. Apart from the funding instrument, referred to in the above, the current quality stimulation program (Quality agendas, developed by the Ministry of Education, since 2007) targeted at basic subjects, could be seen as a possible second favourable systemic factor. As a matter of fact it cannot be excluded that some of the more favourable outcomes with respect to equity are caused by "spillover" from other than the policies specifically dedicated to enhance equity. Apart from the quality agendas, subsequent initiatives to stimulate "inclusive education" might have had a favourable impact, based on increased capacity of teachers to differentiate their teaching. The same might be expected of the stimulation of formative assessment as a part of the policy to stimulate achievement orientation (Dutch: *Opbrengstgericht Werken*).

Groups that manifest lagging behind in a way that appears to be difficult to overcome in the Netherlands are students from Turkish and Moroccan background. The lack of progress in the improvement of their Dutch language skills is particularly worrying.

When it comes to an assessment of the effectiveness of the equity-oriented policies in the Netherlands, there is a striking consensus among all evaluators and reviewers of these policies. They invariably point at the lack of coherence, clear planning frameworks and limited evaluability of the way schools go about using extra funding and working towards the rather general policy objectives (Scheerens 1987; Driessen and Mulder 1999; Mulder and Meijnen 2013). The implicit message is that school autonomy has long gone over the edge in the Netherlands and is preventing policies that are effective and efficient (see the reference to the implications of autonomy 1, 2 and 3 in the above). In the most recent evaluation study, Mulder and Meijnen (2013) are very explicit in their recommendations to have clearer targets from the centre, more explicitly planned programs, stricter accountability requirements and better conditions for program evaluation. The inefficiency in equity-oriented policy is part of a larger syndrome in Dutch education, in which innovation and reform are framed to be "bottom up" see evaluative comments about this approach in earlier sections. Despite of recommendations in practically all evaluation studies, the counsel to make better use of evidence-based comprehensive school reform programs has never been followed up in a consistent way, so far (although there is a very recent initiative to implement the "Success for All" program in the Netherlands).

Proposals for Reform and Specific Measures

I would endorse the specific and concrete proposals for reform and improvement made by Mulder and Meijnen (2013, pp. 1V and V). Some of the main recommendations were already cited in the above:

(a) More central control over the pre-school (VVE) policy implementation.
(b) Stricter accountability on the way schools and school boards apply the extra funding.
(c) Re-formulation of the "weight regulation", to target more funding to those groups most in need, which are young children lagging behind in Dutch language acquisition. A specially developed language test is recommended to identify those most in need.
(d) A research study to try and find out why Dutch, low SES at-risk students, are not improving.
(e) Given the fact that average children outside the target groups of the equity policies are improving, while the target group students do not advance or advance more slowly, it is recommended to rethink the targets of the equity-oriented policies. More long-term policy should be guided by these reformulated targets.
(f) More central guidance in creating conditions for experimental evaluation studies, and continuation of the current monitoring studies.

As much as these recommendations are justified, and deserve support, there is no reason to be optimistic about their implementation. The tradition of "Freedom of Education" has embraced more modern pleas for decentralization and school autonomy. This has a specific interpretation in the resentment against "state pedagogy", and more centrally controlled educational policy. Despite the fairly recent quality agendas and action plans for better performance and some support for evidence-based improvement and innovation, educational governance in the Netherlands depends on compromises with the powerful educational organizations, employers of education and trade unions, partly concentrated now in the councils for primary, secondary, vocational and higher education. Educational support organizations and education consultants are not steered by the centre, but work for individual schools, school boards and clusters of schools and have vested interest in this structure, sometimes mockingly described as the "political educational complex". A factor that is not helpful in bringing about more central control and guidance is the circumstance that in recent years governments have changed frequently, leading to much discontinuity in specific conditions, also as far as the equity-oriented policies are concerned (Scheerens 2013; Nusche et al. 2014; OECD 2014a).

Conclusions About Equity Enhancing Educational Policies in the Netherlands

1. Since 1974, Dutch education policy has invested substantially in compensatory policies to enhance the position of disadvantaged learners. Special funding and other arrangements, as described in this report, have addressed the most important equity challenges, namely the enduring lagging behind of students

from low SES background and (non-western) cultural minorities. These equity-oriented policies can be seen as consistent with labour market policies that seek to improve conditions for employability of people at the margin of the labour market (see EU's Council issues country-specific recommendations on economic and employment policies for the Netherlands; EU 2013). The policies in question are heavily centred on pre-school and primary education (Dutch VVE).

2. Results of monitoring and evaluation provide much relevant information about the policies in question reaching their general targets in terms of improved equity in educational outcomes. Generally, positive outcomes are the relatively high performance of Dutch students in the low tail of the national performance distribution on international assessment tests, and the fact that SES dependency of educational performance is moderate (close to the OECD average), despite the heavily tracked secondary school system and generally selective nature of the Dutch school system. Due to the large autonomy of schools, knowledge about the implementation of equity-oriented strategies is extremely limited and program evaluations have not been able to attribute outcomes to treatments. Successive analyses and evaluations (and the current one is no exception) have repeatedly recommended that policy frameworks should be made more explicit, more central orchestration in policy implementation and more demanding accountability requirements. Unfortunately, such recommendations are difficult to follow-up given the tradition of "freedom of education", governance build on compromise and corporatism and perhaps somewhat exaggerated school autonomy.
3. As far as future priorities are concerned, continuity in the extra funding of disadvantaged groups is desirable. More efforts should be made to enhance the effectiveness and efficiency of the substantial monetary investments. Stronger coordination between equity-oriented policies and adjacent policy programs targeted at inclusive education and a general improvement in educational quality are also desirable.

Final Discussion

In this final discussion, a cursory look is taken at the performance of the Netherlands in international comparisons. Next, the various case studies on quality and equity-oriented educational policy are compared to various improvement scenarios, which were developed in earlier chapters, to try and understand to what degree the relatively positive outcomes of the Netherlands, in terms of quality and equity, can be explained by these policy scenarios.

Performance of Dutch Education in International Perspective

Table 13.3 demonstrates that Dutch students have consistently achieved scores which are (well) above the international average. The Netherlands generally shows very good performance in these international studies, particularly when compared to other European countries. It achieves a slightly lower position on the global ranking due to the exceptionally high scores achieved by countries such as Japan, Singapore, Korea and Taiwan. The Dutch students' scores for science are generally slight lower than those for reading and mathematics (although this trend was broken in PISA 2012). The figures suggest a (slight) downwards trend in mathematics, science and reading alike (where a higher score in science was an exception in PISA 2012). A study by the Central Planning Bureau established that the position of Dutch students in primary and secondary education on international tests is not that splendid at the top of the score distribution. These results are used in a new emphasis on "excellence" in current educational policy, cf. Van der Steeg et al. (2011).

Table 13.3 The Netherlands' results in TIMSS, PIRLS and PISA surveys

Survey	Mathematics		Science		Reading		Problem solving	
	Score	Position	Score	Position	Score	Position	Score	Position
TIMSS 95-3	493	6 (24)	499	6 (24)				
TIMSS 95-4	577	5 (26)	557	6 (26)				
TIMSS 95-7	516	7 (39)	517	10 (39)				
TIMSS 95-8	541	9 (41)	560	6 (41)				
TIMSS 99-8	540	7 (39)	545	6 (39)				
TIMSS 03-4	540	6 (25)	525	10 (25)				
TIMSS 03-8	536	7 (45)	536	8 (45)				
TIMSS 07-4	535	9 (36)	523	17 (36)				
TIMSS advanced 08-12	552	2 (10)						
TIMSS 11-4	540	12 (50)	531	14 (70)				
PIRLS 01					554	2 (35)		
PIRLS 06					547	12 (45)		
PIRLS 11					546	13 (45)		
PISA 00	564	1 (42)	529	6 (42)	532	3 (42)		
PISA 03	538	4 (40)	521	8 (40)	513	9 (40)	520	12 (40)
PISA 06	531	5 (57)	525	9 (57)	507	10 (57)		
PISA 09	526	11 (66)	522	11 (66)	508	10 (66)		
PISA 12	523	10 (65)	511	15 (65)	522	14 (66)		

The figure behind the year of reference is the grade level that was assessed in TIMSS or PIRLS; PISA results are for 15-year-old students

In the previous section, it was shown that the Netherlands, despite the tracked school system in secondary education, does relatively well on a number of equity indicators.

Tentative Attribution to Improvement Scenarios

Substantively credible and simple scenarios for systemic reform that were addressed in previous chapters are: (a) the **teacher training and recruitment scenario**: selective recruitment of teachers, teacher training providing sufficient support for teacher knowledge, particularly pedagogical content knowledge, all supporting effective teaching, featuring classroom management, support and cognitive activation, as the most proximal stimulant of improved student performance; (b) **the accountability scenario**: setting standards (as in a national core curriculum), developing and applying well-aligned summative tests and evaluations, possibly supported by more minute, well-aligned monitoring systems for formative evaluations, and opportunity to learn and test preparation (content aligned to standards and examinations) as the most proximal condition to stimulate student achievement; (c) **the good governance scenario**: patterns of functional decentralization fitting to the historical and cultural context of the country in question, lean, professional government and school management provisions, including a basic "meta control" orientation, leaving room for distributive and "teacher leadership", facilitating task-oriented teacher cooperation and professional development; (d) **the evidence-based comprehensive reform scenario**: national stimulation of the implementation of evidence-based comprehensive school reform programs, in which school-level aspects of the first three scenarios are integrated.

When considering the Dutch quality-oriented educational policy, the emphasis should be on two of these scenarios, the good governance scenario, which is based on school autonomy, and the accountability scenario, about evaluation and assessment provisions. These two facets seem most prominent in distinguishing the Dutch educational policy context. Although there is definitely a lot of attention for the teacher recruitment and training scenario in current educational policy, it does not appear to differ strongly from developments in other industrialized countries. Possibilities for evidence-based comprehensive reform in the Netherlands are discussed in relationship to the governance structure and school autonomy.

Autonomy is a part of the DNA of the educational system in the Netherlands. It is based on the principle of freedom of education, which gave denominational schools equal rights as public schools. Educational organizations and "pressure" groups, such as unions of school governors and trade unions, were organized according to "denominational pillars"; a catholic pillar, a protestant pillar and a neutral pillar. Together, these organizations and unions formed a strong middle level of influence and power, which has been described as having corporatist characteristics (Leune 1983). This strong countervailing power of educational

organizations exists until today in a modernized form, where the pillar organizations have been replaced by the primary and secondary school councils (PO and VO Councils). This structure expresses a general aversion of "state pedagogy", and has particular implications for quality-oriented policy and the role of the educational support structure. Reform and innovation policy has to be "bottom up", and it is often stated that schools are the units responsible for quality improvement. Educational support organizations, even when working within the general frame of policy targets agreed at the centre between the government and the educational organizations, do **not** function as a technostructure that facilitates implementation of national policy, but they are controlled by the autonomous schools, or networks of schools that share the same school board. This approach to educational reform and school improvement has only been challenged once, in the recent history, in the form of the "constructive educational policy" of the social democrat Minister of Education, Van Kemenade, in the 1970s. According to this policy, reform was initiated from the centre and reform programs had to be based on scientific data. Innovation programs were tried out as relatively small-scale field experiments, which would only be implemented at the national level, after scientific evaluation had shown them to be effective. The best known example of the innovation experiments was the middle school experiment, an attempt to integrate secondary school categories into one comprehensive school. The constructive educational policy had to be abandoned for political reasons (a change in government, leading to a take over by the Christian democrat and conservative liberal parties), but also because of "technical" problems. The development of innovatory programs did not get off the ground, and the program evaluations were boycotted by the schools. The situation in the 1970s has much in common with the way current quality-oriented policy is faring, with a strong emphasis on bottom-up reform, and more marginal attempts at evidence-based policy. Like in the 1970s, evaluation research, particularly attempts at carrying out randomized field trials, suffers from lack of cooperation from the autonomous schools. Against this background, educational policy in the Netherlands, calling for more autonomy by successive coalition governments, has a large degree of being symbolic. The school system is autonomous from day one, and stressing autonomy and bottom-up reform is more like restating the status quo in the educational province.

The coincidence that the Netherlands has high school autonomy as well as fairly high educational performance in international comparisons has prompted the belief that the Dutch educational system is doing quite well, because schools are autonomous. Evidence that was presented in previous chapters has indicated that generally, across countries, there is no convincing evidence that autonomy can be seen as a cause of high educational performance. This means that one should be prudent in drawing conclusions about autonomy causing high educational performance. Doubts about the effectiveness of school autonomy are based on two kinds of arguments: (a) the supposed effectiveness-enhancing mechanisms for linking autonomy and performance are less convincing and (b) autonomy is seen as effectively blocking other "real" effectiveness-enhancing mechanisms, like

standard-based reforms and assessment. Arguments in favour of high autonomy being effective are more flexibility to adapt to the local context, more flexibility in meeting specific demands of the students in a school, more room for innovation and creativity at school level and less administration and bureaucracy. Arguments about autonomy "blocking" other effectiveness-enhancing mechanisms relate to phenomena like implementation failures of externally structured innovations, resistance against external school evaluation and national assessment programs, predominance of self-centred idiosyncratic interests, vulnerability to inputs from consultancy and support organizations promoting organizational development because this is their core business, and risk of goal displacement when popular treatments like "peer learning of teachers" and "professional learning communities" are disconnected from a focus on effective teaching. As it comes to applying these lines of argumentation to the situation in the Netherlands, one wonders why the system is doing relatively well, despite the institutionalization of autonomy. The alleged positive mechanisms that would supposedly enhance student performance because of high school autonomy are less powerful and unique. Even when schools are less, autonomous teachers can still be very autonomous in adapting to the needs of the group of students that they are teaching. The idea of school autonomy fostering high creativity does not see the reality of a sector where the large majority is coping and surviving, not in a position to be constantly involved in school organizational development. In addition, in the Netherlands school autonomy constrains evidence-based educational policy. Because of the strongly felt aversion against anything that just has a flavour of state pedagogy, evidence-based, externally induced innovation is likely to be avoided and actively blocked. So much even that hardly anyone is still making a case for the implementation of evidence-based programs. A positive exception is the introduction of "Success for All" in the province of Groningen, supported by the University of Groningen. Finally, summative program evaluations have been effectively resisted by autonomous schools, since they were first proposed in the 1970s. The experience with the educational priority policy program illustrates that autonomy does not automatically lead to schools choosing the most effective strategies. A curious result is that despite of the fact that program evaluations have rarely found any positive evidence supporting the policy, the Netherlands still does surprisingly well on equity indicators for an educational system that is strongly tracked.

Examinations and high stakes testing form one area where the Dutch system is not so autonomous. Some elements in the positive causal chain in the "accountability" scenario seem to be clear in evidence in the Netherlands: In the wake of high stakes summative tests and examinations elaborate pupil monitoring systems are used in primary, and to a lesser degree, also in lower secondary schools. Next, the national policy of stimulating "achievement oriented work" leans heavily on formative evaluation, by means of student monitoring systems. There is no research-based evidence on tests and examinations being used as guidelines for the selection of content in text books and teaching. Still it is public knowledge that schools frequently practice test and examination preparation, by means of

illustrative items. Given the quality of the summative tests and examinations (use of item banks, for example), I would tend to see these practices as legitimate measures to enhance "opportunity to learn", rather than "teaching to the test". When it comes to find explanations for the relatively strong performance of the Netherlands on international assessments, the "accountability scenario" is a strong hypothetical candidate.

There is some empirical support for the hypothesis that the *combination* of high autonomy on processes and central control on outcomes is effective (Woessmann et al. 2009; OECD 2010, 2014a). Such a combination confirms to the principles of "new public management", where the advantages of strong professional autonomy on work processes and firm control on outcomes are expected to provide the best results. Although in general terms this combination is very much in evidence in the way the Dutch educational system is structured, I would hesitate to accept it at face value. In my view, the way autonomy plays out has important dysfunctional elements, see the arguments presented in the above. When writing these lines (July, 2015), I am looking at an invitation by the Ministry of Education to contribute to a briefing of OECD examiners, who will do a review of the Dutch educational system. The issue that I am asked to comment on is "how the outcomes of Dutch education can be improved and moved from 'good to great'". I find it somewhat ironic that the "process" characteristics of Dutch education already confirm strongly to the conditions of greatness that are recommended in the McKinsey report to which we owe these levels of excellence. Recommended are high school autonomy, a strong midfield, including support organizations and much attention to secondary organizational functions of schools like "peer learning" of teachers, teachers as researchers and networking with other schools. All these structural arrangements are in place and the organizational development facets predominate (see the example of "Schools have the initiative"). So according to this reasoning the Netherlands should be considered well on its way to reach the outcome standards of greatness, which would come down to an increase of at least 10 points on the universal scale score, designed by McKinsey (Mourshed et al. 2010, 21). Previous chapters of this book showed that progress caused by the mostly considered factors in systemic reform and school improvement should be expected to come in "small steps" only. As it comes to incremental progress, the scenario of accountability and enhanced opportunity to learn has more credibility than the strategies that capitalize on autonomy and organization development, which, in the Dutch context, might even be considered as a threat to maintaining the current performance level. OECD's recommendation (Nusche et al. 2014) to further boost the evaluation and assessment framework by means of a system of long-term learning goals, to which all specific evaluation activities should be aligned, makes a lot of sense but will be difficult to realize given the strong tradition of autonomy and the aversion against "state pedagogy".

Annex: Schematic Overview of Evaluation and Assessment Provisions in Dutch Education

Overview A: Type of Evaluation, Short Description, Formal Responsibility and Implementation and Use

System evaluation

Type of evaluation	Short description	Formal responsibility	Implementation and use
Policy and program evaluations	Evaluation of educational policies and programs	Minister of Education, in one occasion Parliament	Strong resistance from the field against early program evaluations. Little evidence on actual use
PPON	Periodic national assessment primary schools	Central Test Agency, CITO	Relatively low profile
Cohort studies	Achievement and attainment indicators of cohorts of primary and secondary school students	Joint responsibility of the Ministry of Education and the Foundation for Scientific research, NWO	Question marks with respect to use by education policy planners and schools
Annual report Inspectorate	Comprehensive report on the state of education	The Inspectorate of Education	Relatively high profile for policy use. Modest press coverage
Monitors	Partial effect and evaluation studies contracted out by the Ministry of education, some of them longitudinal (monitors)	The Ministry of Education	Extensive information, no clear evidence about synthesis and policy use
Key data, "trends in beeld"	Comprehensive annual reports containing key data and indicators	The Ministry of Education	Appear to have high potential for policy use, given active dissemination and user friendly formats
International indicators and international assessments	Dutch participation in IEA, EU and OECD studies	The Ministry of Education	Have obtained high profile in public debate on education concerning the quality of education

School assessment

Type of evaluation	Short description	Formal responsibility	Implementation and use
School self-evaluation	Internal quality care by schools	Schools	Hampering implementation, substantial underutilization
School inspection	Systematic school supervision, using structured formats and check lists	Inspectorate of Education	No implementation problems, schools have a positive attitude to inspections
Quality cards	User friendly set of key indicators on school functioning to inform general public and parents. Recently reduced to an indication of the inspection regime a school has to follows, which is indicative on good or failing performance	Inspectorate of Education	Disappointing use by parents for purposes of school choice
Windows for accountability	Information dossiers on each school, consisting on centrally delivered quantitative indicators and qualitative indicators provided by schools	A new foundation resorting under the Councils for Primary and Secondary Education, as of 2012	No use and impact information available as yet

Teacher appraisal

Type of evaluation	Short description	Formal responsibility	Implementation and use
New initiative Inspectorate to appraise teachers	Inspection of personnel policy of schools and the quality of teaching in a school; classroom observations in a national sample of schools	The Inspectorate of Education	Results are published in the Annual Inspection Report
Within school teacher supervision	Individual teacher appraisal by school leadership and governance	The competent authorities of the school	No systematic information available

Chapter 6 Student assessment

Type of evaluation	Short description	Formal responsibility	Implementation and use
Examinations	Formal assessments at the end of secondary schools for purposes of individual certification	The Ministry of Education, with delegated responsibility to the CVE and CITO. Schools, monitored by the Inspectorate are responsible for the internal school examination	Implementation is obligatory. Use and application is straightforward
CITO test	The CITO test is a school leaver test at primary school level, used by 85 % of schools	Schools are responsible for taking part. CITO takes care of technical aspects	The test is used in supporting students' choice of a specific secondary school track. In aggregated form, use for school and system-level evaluation
CITO LVS	A pupil monitoring system for primary schools, all grades and broad coverage of subjects	Schools are responsible for taking part, i.e. they buy into the system. CITO takes care of technical aspects	Tests are used for didactic diagnosis and formative student assessment. In addition aggregated data are sometimes used for school self-evaluation. Actual use by schools is still far from optimal

Overview B: Type of Evaluation, Short Description, Main Instruments and Periodicity

System evaluation

Type of evaluation	Short description	Main instruments	Periodicity
Policy and program evaluations	Evaluation of educational policies and programs	(Failed) attempts at quasi-experimental designs in 1970s Ex post facto research in evaluations 1980s, 1990s and first decade of 2000	1970 Evaluation of Innovation programs 1980–1990 Retroactive evaluations secondary and primary education 2008 Parliamentary Committee Educational Innovation (Dijsselbloem)

(continued)

(continued)

Type of evaluation	Short description	Main instruments	Periodicity
PPON	Periodic national assessment primary schools	Achievement tests	Since 1987 Mathematics/Arithmetic: 1999/2000, 2005, 2012, 2013 Dutch language: 2002 and 2005, 2011, 2013 Reading skills: 2007, 2008, 2011 World orientation, diverse subjects; 9 reports between 2003 and 2011 Non-cognitive skills, social outcomes, citizenship, 2011, 2012
Cohort studies	Achievement and attainment indicators of cohorts of primary and secondary school students	Achievement tests in mathematics/arithmetic and language; educational attainment data (progress, graduation); questionnaires to obtain school background information	1977 SMVO-cohort 1982 SLVO-cohort VOCL—cohort: 1989, 1993, and 1999/2000 PRIMA—cohort: 1988, 1994/95, 2000/01, 2002/03, 2004/05 2007 COOL—cohort, integration primary and secondary cohorts
Annual report Inspectorate	Comprehensive report on the state of education	Secondary analyses Questionnaires Direct observations by Inspectors	Annually, since 1801
Monitors	Partial effect and evaluation studies contracted out by the Ministry of education, some of them longitudinal	Varied methodology; mostly based on questionnaires; panel studies; last 5 years attempts at randomized control studies	Permanently, yearly updating
Key data, “trends in beeld”	Comprehensive annual reports containing key data and indicators	Education statistics	Since 1996, on an annual basis
International indicators and international assessments	Dutch participation in IEA, EU and OECD studies	Internationally comparative assessment tests; school and teacher questionnaires	TIMSS since 1995 PISA, since 2000

School assessment

Type of evaluation	Short description	Main instruments	Periodicity
School self-evaluation	Internal quality care by schools	Check lists, questionnaires; narrative reports, self-assessments	Strong stimulation in the period between 1999 and 2006 Continued press caused by the Inspectorate observing "quality care" as a core quality facet
School inspection	Systematic school supervision, using structured formats and check lists	Explicit evaluation frameworks and structured observations, since about 1986	Permanently, according to selection schemes of schools that have changed over time. Currently each school is inspected at least once every four years
Quality cards	User friendly set of key indicators on school functioning to inform general public and parents. Recently reduced to an indication of the inspection regime a school has to follows, which is indicative on good or failing performance		Quality cards from the Inspectorate came into existence in 1998 for secondary schools and 2003 for primary schools. In 2007 replaced by much reduced cards on the "supervision arrangements" for schools; related to current proportional inspection. In addition so-called "attainment cards", for secondary schools came into being
Windows for accountability	Information dossiers on each school, consisting on centrally delivered quantitative indicators and qualitative indicators provided by schools	Administrative data available at the national level for the "central" Indicators. Local indicators are based on information provided by schools, e.g. on student and parent satisfaction	Windows for secondary schools is operational since 2010. Windows for primary schools started in 2012

Teacher appraisal

Type of evaluation	Short description	Main instruments	Periodicity
New initiative Inspectorate to appraise teachers	Inspection of personnel policy of schools and the quality of teaching in a school; classroom observations in a national sample of schools	Adapted school evaluation framework of the Inspectorate, 2013 revised supervision framework Structured observation by Inspectorates	Part of the regular schedule for school inspections, including proportional inspection and basic inspection of all schools, every four years
Within school teacher supervision	Individual teacher appraisal by school leadership and governance	Criteria to evaluate teacher competencies by school boards and school leaders. National register of qualified teachers	Since 2006. Teacher register since October 2011

Student assessment

Type of evaluation	Short description	Main instruments	Periodicity
examinations	Formal assessments at the end of secondary schools for purposes of individual certification	Central and school examinations consist of multiple choice tests and open questions.	Each year
CITO test	The CITO test is a school leaver test at primary school level, used by 85 % of schools	Standardized achievement test, multiple choice (language, mathematics/arithmetic, study skills)	Each year
CITO LVS	A pupil monitoring system for primary schools, all grades and broad coverage of subjects	Standardized tests for longitudinal assessment in all main subject matter areas in all grades of the primary school (see Table 6.1, country background report	Test taking as part of the LVS typically occurs twice, during a school year

References

Bronneman-Helmers, R., the Social Cultural Planning Bureau. (2011). *Overheid en onderwijsbestel Beleidsvorming rond het Nederlandse onderwijsstelsel, (1990–2010*). The Hague: SCP-publication 2011-31.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets and American schools*. Washington, D.C.: Brookings Institute.

De Boer, H., Bosker R.J., Werf MPC vd. (2007). De gevolgen van onder—en overadvisering. In I. v. h. Onderwijs (Ed.), *Inspectie van het Onderwijs, Onderadvisering in beeld* (pp. 83–92). Utrecht: Inspectie van het Onderwijs.

De Wolf, I., & Verkroost, J. (2011). Evaluatie van de theorie en praktijk van het nieuwe onderwijstoezicht. (Evaluation of theory and practice of the new proportional inspection). *Tijdschrift voor Toezicht*, *2*(2), 7–24.

Driessen, G., (2011) Onderwijsadvisering van allochtone leerlingen. Raboud Universiteit, ITS http://www.onderwijsinspectie.nl/actueel/publicaties/Onderadvisering+van+allochtone+leerlingen.html

Driessen, G. (2013) De bestrijding van Onderwijsachterstanden (Fighting educational inequality. A review of effective approaches). Nijmegen: ITS internal report.

Driessen, G., & Mulder, L. (1999). The enhancement of educational opportunities of disadvantaged children. In R. J. Bosker, B. P. M. Creemers, & S. Stringfield (Eds.), *Enhancing educational excellence, equity and efficiency*. Dordrecht: Kluwer.

Education Council (*Onderwijs Raad*). (2012). Geregelde ruimte (p. 76). Den Haag.

Educational Council (*Onderwijsraad)*. (2013). *A narrow view on educational quality*. The Hague: Educational Coucnil.

EFEE. (2012). *School leadership and governance. Lifelong learning*. Brussels: EFEE.

ETUCE (2012) School leadership. Budapest: ETUCE 27 November.

European Commission (2013) Council issues country-specific recommendations on economic and employment policies for the Netherlands. http://register.consilium.europa.eu/pdf/en/13/st12/st12073.en13.pdf

Frissen, P. H. A., Hofman, R. H., & Peeters, P. (2013). *Het ongemak van autonomie. Onderwijsbeleid tussen vrijheid en verantwoording.*. Groningen: GION.

Hooge, E., & Honingh, M. (2014). Are school boards aware of the quality of their school? *Educational management, administration and leadership*. Available online at SagePub.

Leune, J. M. G. (1983). *De reikwijdte van het onderwijsaanbod in het funderend onderwijs: Referaat ten behoeve van het RION-congres op 28 en 29 maart 1983 te Groningen over "De kwaliteit van het onderwijs"*. Rotterdam: EUR.

Leune, J. M. G. (2007). *Verstandig onderwijsbeleid*. Garant: Antwerpen; Apeldoorn.

Luyten, H. (2011). The quality of education in the Netherlands, as expressed by achievement and attainment indicators. In J. Scheerens, H. Luyten, & J. Van Ravens (Eds.) *Perspectives on educational quality. Illustrative outcomes on primary and secondary schooling in the Netherlands.* Heidelberg, New York: Springer.

Luyten, H., Scheerens, J., Visscher A., Maslowski., R., Witziers, B., & Steen, R.. (2005). *School factors related to quality and equity. Results from PISA 2000.* Paris: OECD.

Ministry of Education. (2007a). *Scholen voor morgen*. Kwaliteitsagenda PO. Den Haag: Ministerie OCW.

Ministry of Education. (2007b). *Tekenen voor Kwaliteit*. Kwaliteitsagenda Voortgezet Onderwijs. Den Haag: Ministerie OCW.

Ministry of Education. (2008). *Werken aan vakmanschap*. Strategische agenda Beroepsonderwijs en Volwasseneneducatie 2008–2011. Den Haag: Ministerie OCW.

Ministry of Education. (2012). *Stand van Zaken Zicht op Kwaltiteit*. The Hague: Internal report.

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. McKinsey and Company.

Mulder, L., & Meijnen, W. (2013) Onderwijsachterstanden in de BOPO-periode, 2009–2012. Een review. (Inequality in primary education in the period 2009–2011. A review study. Nijmegen: ITS.

Nusche, D., Braun, H., Halasz, G., & Santiago, P. (2014). *OECD reviews of evaluation and assessment in education: Netherlands 2014*. Paris: OECD Publishing.

OECD. (1996). *Education at a glance*. Paris: OECD.

OECD. (2008). *Education at a glance*. Paris: OECD.

OECD. (2010). *PISA 2009 results volume I, what students know and can do: student performance in reading, mathematics and science*. Paris: OECD Publishing.

OECD. (2012a). *Equity and quality in education: Supporting disadvantaged students and schools*. Paris: OECD.

OECD. (2012b). *Education at a glance*. Paris: OECD.

OECD (2014a) *PISA 2012 results, volume IV, what makes schools successful? Resources, policies and practices.* Paris: OECD Publishing.

OECD. (2014b). *PISA 2012 results: Excellence through equity* (Vol. II). Paris: OECD.

Scheerens, J. (1987). *Enhancing opportunities for disadvantaged learners. A review of Dutch research on compensatory education and educational development policy*. Amsterdam: North Holland Publishing Company.

Scheerens, J. (2013). *Educational evaluation and assessment in The Netherlands. Country background report for the OECD study on educational evaluation and assessment. Addendum*. Enschede: University of Twente.

Scheerens, J. (2014). *Country report on The Netherlands, concerning equity oriented educational policies and programs. Network of experts on social aspect of education and training, NESET*. Cardiff: University of Cardiff.

Scheerens, J., & Doolaard, S. (2013). *Review studie Onderwijskwaliteit PO*. Groningen: GION.

Scheerens, J., Ehren, M., Sleegers, P., & de Leeuw, R. (2012). *Educational evaluation and assesment in the Netherlands. Country bacground report for the OECD study on evaluation and assessment frameworks for improving school outcomes*. Enschede: University of Twente.

Scheerens, J., Luyten, H., & Glas, C.A.W. (2013). *Trends in equity. Data based on PISA 2009.* Unpublished manuscript, University of Twente.

Scheerens, J., Luyten, H., & Van Ravens, J. (2011). *Perspectives on educational quality. Illustrative outcomes on primary and secondary education in the Netherlands.* Research Briefs, Springer.

Van der Steeg, M., Vermeer, N., & Lanser, D. (2011). Niveau onderwijs daalt. Vooral beste leerlingen blijven achter *Nederlandse onderwijsprestaties in perspectief*. (The level of education is going down. The highest scoring pupils are lagging behind) The Hague: CPB.

Van der Werf, M. P. C. (1995). *The educational priority policy in the Netherlands.* The Hague: SVO. (ISBN 90-6472-173-4).

Van Ravens, J. (2011) The selectivity and vocational orientation of education systems. In J. Scheerens, H. Luyten, & J. Van Ravens (Eds.) *Perspectives on educational quality. Illustrative outcomes on primary and secondary schooling in the Netherlands.* Heidelberg, New York: Springer.

Visscher, A. J., & Ehren, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken*. Enschede: Universiteit Twente, Vakgroep Onderwijsorganisatie en-management.

Woessmann, L., Luedemann, E., Schuetz, G., & West, M. R. (2009). *School accountability, autonomy and choice around the world*. Cheltenham, UK: Edward Elgar.

Epilogue

In the epilogue to this book attention is given to the limitations of the work, to key issues identified and their possible implications for further research and analysis.

Limitations

The conceptual models that were used in Part I, to provide an orderly description of the way key factors are expected to work, suggest a fairly modest level of complexity. Yet, the bulk of primary research and meta-analyses that were referred to study only the more basic relationships between malleable conditions and outcome measures. In that way most of the empirical research does not do justice to even this limited complexity of the models. Although the study of mediated and indirect effects is increasing in recent years, over the whole period that was reviewed this represents only a minority of studies. Meta-analyses of indirect effect models hardly exist.[1] Despite the general recognition of the validity of contingency theory for educational effectiveness, formal tests of contingency hypotheses, in the sense of interactions between effectiveness enhancing and context characteristics are rarely carried out. Despite the recognition that frequently addressed school factors, such as "high expectations" and "collective efficacy" have a reciprocal relationship with performance, feedback loops and non-recursive models are hardly ever studied. And, to give a final illustration, very few studies have addressed the influence of system-level malleable conditions on student achievement, mediated by school-level variables. To this should be added that more complex models than the ones described in Part I might be required, in which nonlinearity and vicious or virtuous cycles of events are recognized. The knowledge on educational effectiveness presented in this book depends very much on "the robust beauty of improperly specified linear models", to quote Lee Sechrest.

[1]Our study on School Leadership Effects (Hendriks and Scheerens 2013) may be one of the first attempts.

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8

The knowledge base on educational effectiveness is loosely defined as the whole of available research evidence, areas of consensus in the relevant scientific community, as well as theoretical understanding. When it was attempted to review this "knowledge base" it should be evident that, literally speaking, this is an impossible task, if only because its borders are ill-defined and it is a moving target, to which new studies and meta-analyses are added every day. In earlier references the phrase "providing glimpses of the knowledge base" was used to put the effort into a certain perspective. Moreover Part II, in discussing empirical research evidence, heavily depends on our own reviews and meta-analyses and cannot claim representativeness.

Key Issues and Implications for Further Research and Analysis

The key message from this book is that malleability in education is limited. This outlook is strongly prompted by our own research results, but is also based on the lack of consensus across meta-analyses carried out by others. The field faces the nagging challenge of "unit effects", country, school and teacher effects, that are only partially explained by the modelled malleable variables, despite their strong plausibility. Ongoing search for strong treatments and further sophistication in research approaches in doing more justice to the complexity of the object of study, already mentioned in the above, are logical reactions. In addition more attention should be given to alternative explanations, based on the influence of contextual characteristics and "given" characteristics of the units in question, educational systems, school communities, school leaders, teachers and students. For example, at the level of educational systems, the weak empirical support for most of the system-level levers of educational reform, such as autonomy, school choice and to a certain extent even accountability, could be taken as a sign to look for the influence of more fundamental characteristics of national cultures, described with assets like achievement orientation, discipline, trust, the valuing of education and innovatory potential (cf. the work of Hofsted). An interesting illustration is the study by Sahlgren (not to be confused with Sahlberg), who offers an explanation of Finland's educational performance, in which recent success and current decline are associated with a history of centralistic governance and traditional teaching. In his analysis, the author speaks of "the iron cage of history" and sees the current preference for a less authoritative schooling culture as a basis for an oncoming decline (Sahlgren 2015). At school level composition effects of the student body as well as the teaching staff form a very interesting research area. Next, personality characteristics of teachers and principals should perhaps be reconsidered as sources of explanation in educational effectiveness, particularly also within the context of selection and recruitment procedures.

System effectiveness and systemic reform are important new areas in educational effectiveness, very much stimulated by international comparative assessment studies, and country reviews that have made extensive use of their results. In the

interpretation of the international research results there appears to be a revival in recognizing the role of the centre, next to a continuation of decentralization policies, and emphasis on school and teacher autonomy. The term "alignment" between levels and between policy areas (such as examinations and curricula) is frequently used. Still there are important trade-offs between centralized coordination and autonomy at lower levels. In some countries decentralization and autonomy have reached a level where the steering potential of the centre has been strongly diminished (see the case study of the Netherlands, and, to some extent, the OECD (2014) country review of Sweden). The steering capacity of national governments deserves a more prominent place on the research agenda on system-level educational effectiveness.

The rationale for this book has been the idea that, in order to learn about educational effectiveness, we should better understand ineffectiveness. In educational effectiveness research and application there is a strong expectation for positive "how to do" knowledge. In this context dissonant information that challenges the claims of the predominant positive theory is particularly important, (although it is no more than the basic mission of research). In essence a critical attitude is constructive as it may forewarn against inefficient investments in reform. Recent manifestations of this critical attitude are the study by Sahlgren that was cited in the above and articles by Marks (2015a, b) on the interpretation of composition effects and the low stability and consistency of school effects.

References

Hendriks, M. A., & Scheerens, J. (2013). School leadership effects revisited: A review of empirical studies guided by indirect-effect models. *School Leadership & Management*, *33*(4), 373–394. http://doc.utwente.nl/86793/.

Marks, G. N. (2015a). Are school-SES effects statistical artefacts? Evidence from longitudinal population data. *Oxford Review of Education, 41*(1), 122–144.

Marks, G. N. (2015b). The size, stability, and consistency of school effects: Evidence from Victoria. *School Effectiveness and School Improvement, 26*(3), 397–414.

OECD. (2014). *Improving schools in Sweden. OECD policy reviews*. Paris: OECD Publishing.

Sahlgren, G. H. (2015). *Real finish lessons. The true story of an education superpower*. Center for Policy Studies. Surrey.

Index

J. Scheerens, *Educational Effectiveness and Ineffectiveness*,
DOI 10.1007/978-94-017-7459-8