Review of Educational Research
December 2007, Vol. 77, No. 4, pp. 454–499
DOI: 10.3102/0034654307310317

Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results

Tina Seidel
Leibniz-Institute for Science Education at the University of Kiel and Friedrich Schiller University Jena, Germany

Richard J. Shavelson
Stanford University

This meta-analysis summarizes teaching effectiveness studies of the past decade and investigates the role of theory and research design in disentangling results. Compared to past analyses based on the process–product model, a framework based on cognitive models of teaching and learning proved useful in analyzing studies and accounting for variations in effect sizes. Although the effects of teaching on student learning were diverse and complex, they were fairly systematic. The authors found the largest effects for domain-specific components of teaching—teaching most proximal to executive processes of learning. By taking into account research design, the authors further disentangled meta-analytic findings. For example, domain-specific teaching components were mainly studied with quasi-experimental or experimental designs. Finally, correlational survey studies dominated teaching effectiveness studies in the past decade but proved to be more distal from the teaching–learning process.

Keywords: research design, meta-analysis, models of teaching and learning, process–product model, student learning, teaching effectiveness.

This meta-analysis summarizes studies of the past decade that have investigated the effects of teaching on student learning. Thus, the focus is on the components of teaching in classrooms and their effects on student learning. One reason for conducting this meta-analysis in this time period was that recent models of teaching and learning should have influenced the findings of empirical studies. Consequently, prior meta-analyses needed to be updated. The second and equally important reason was to use meta-analysis to address potentially conflicting findings among studies.

Accordingly, we based the meta-analysis on recent models of teaching and learning and distinguished among research designs. Thus, by being explicit about current and past models and evaluating evidence methodologically, we sought to provide in-depth information about teaching effects on student learning. By taking this tack, we also sought to enhance future teaching effectiveness research both conceptually and methodologically.

We are aware that studies of the effects of teaching on student learning are diverse and that a meta-analysis in this area of research has to summarize many types of instructional approaches. Decisions about how to categorize diverse types of teaching acts had to be made; an analysis of each individual technique was not possible. The goal of the meta-analysis, however, was not to endorse any particular category of teaching approach but to summarize the state of research after a certain period of time in a manner useful to researchers and policy makers.

Recent models of teaching and learning characterize learning as a self-regulated and constructive process (Bransford, Brown, & Cocking, 2000; Collins, Brown, & Newman, 1989). Moreover, this characterization has stimulated a substantial number of empirical studies of the effects of teaching on student learning, using a variety of methods. By making explicit the common components in these models, we seek to focus researchers' attention on components of learning and on teaching as a process of creating and fostering learning environments in which students are supported in activities that have a good chance of improving learning.

We examine the research in three ways. The first way recognizes that the effects of teaching on student learning can be diverse. Some teaching components might have an effect on students' cognitive growth, others on students' motivational development, and other learning components might affect learning processes. Moreover, some components might show short-term effects on learning processes, whereas others might show long-term effects on motivation and achievement. Thus, the meta-analysis investigated the effects of teaching on three types of outcome measures found in research differentiating between the effects of teaching on cognitive and on motivational–affective student outcomes (Gruehn, 1995): learning process outcomes, cognitive outcomes, and motivational–affective outcomes (Shuell, 1993, 1996; Snow, Frederico, & Montague, 1980).

The second way we examine current research is by categorizing studies into one or another model of teaching and learning and by investigating the effects of model components on student outcomes. The first model, the process–product model, has served as the basis of previous meta-analyses. We apply this first model in our meta-analysis to provide continuity in findings on teaching effectiveness research. However, this model is limited in light of research conducted in the past decade. Models of teaching and learning have changed to emphasize cognitive components and have been refined. Thus, a second model that is based on current conceptualizations of teaching and learning, a cognitive learning process model, is used. We chose this cognitive conceptualization of teaching and learning instead of alternative models that stress the social and situative perspective (Floden, 2001) because the body of studies with a background in a social and situative perspective was too sparse during this time period. Although we speak of a cognitive model, we nevertheless conceptualize teaching and learning jointly, and we consider context (Shuell, 1993, 1996). Thus, our model is one of teaching and learning in context, in which elements of teaching are integrated into components of learning.

The third way we examine the research is by accounting for the role that research design plays in investigating teaching effectiveness. We ask whether effect sizes vary systematically with different designs.

In Part I, we present a framework that accounts for models of teaching and learning. In doing so, we briefly review and summarize the field of teaching effectiveness. In Part II, we take up methodological considerations and challenges in teaching effectiveness research. Part III focuses on the research questions, methods, and results of the meta-analysis. In Part IV, we summarize our findings and draw conclusions with regard to the future developments of teaching effectiveness research.

Before proceeding, we need to clarify some vocabulary. We speak of *teaching variables* when we refer to single teaching acts or behaviors as stated in published research. We refer to *teaching components* when we classify and interpret teaching variables according to theoretical models. We speak of *teaching effects* or *teaching effectiveness* when referring to the effects of teaching on student learning. Finally, we use the term *models of teaching and learning* as shorthand for the effects of teaching on learning.

I. Models of Teaching and Learning

Traditional Models in Meta-Analyses of Teaching Effectiveness

Early teaching effectiveness research hypothesized that certain teaching acts and conditions would affect student outcomes. The basic process–product model distinguished product, process, presage, and context variables (Dunkin & Biddle, 1974). Product variables referred to a variety of student outcomes such as student achievement or attitudes toward learning. Process variables were defined as the conditions of teaching that enhance student outcomes (products). In this sense, various teaching approaches—such as direct teaching, reinforcement, time on task—served as process variables. Presage variables took into account the characteristics and prior knowledge of students (age, ability, gender). Context variables (such as parental involvement, instructional media) were considered to be factors that influence an effect of teaching on student outcomes.

Almost all reviews and meta-analyses of school, teacher, and teaching effectiveness have been based on the process–product model (Anderson, 2004; Brophy, 2000; Brophy & Good, 1986; Creemers, 1994; Doyle, 1986; Fraser, Walberg, Welch, & Hattie, 1987; Gage, 1963; Gage & Berliner, 1998; Helmke & Weinert, 1997a; Lipsey & Wilson, 1993; Muijs & Reynolds, 2000; Ross, Stringfield, Sanders, & Wright, 2003; Scheerens, 2000; Scheerens & Bosker, 1997; Scheerens, Seidel, Witziers, Hendriks, & Doornekamp, 2005; Shuell, 1993; Shulman, 1986; Snow & Swanson, 1992; Walberg & Paik, 2000; Wang, Haertel, & Walberg, 1993; Wayne & Youngs, 2003). (Of course, the way teaching variables were organized differed among reviews and meta-analyses.) Given the sheer body of reviews and handbook chapters, an integral summary of findings is beyond the scope of this article. However, as an example, we draw on the results of the two most comprehensive recent meta-analyses, those conducted by Scheerens and Bosker (1997) and by Fraser et al. (1987). They reported that a number of teaching process variables (e.g., reinforcement) positively affected student outcomes (see Table 1).

The five teaching components with the highest effect sizes in Fraser et al.'s (1987) meta-analysis were reinforcement, acceleration, reading training, cues and feedback, and science mastery. In Scheerens and Bosker's (1997) meta-analysis,

TABLE 1
Effective teaching components as reported in two previous meta-analyses

Fraser, Walberg, Welch, and Hattie (1987, p. 157)	All outcomes	Scheerens and Bosker (1997, p. 305)	All outcomes
Reinforcement	1.17	Reinforcement	0.58
Acceleration	1.00	Feedback	0.48
Reading training	0.97	Cooperative learning	0.27
Cues and feedback	0.97	Differentiation/adaptive instruction	0.22
Science mastery	0.81		
Cooperative programs	0.76	Time on task	0.19
Reading experiments	0.60	Structured teaching	0.11
Personalized instruction	0.57	Opportunity to learn	0.09
Adaptive instruction	0.45	Homework	0.06
Tutoring	0.40		
Individualized science	0.35		
Higher order questions	0.34		
Diagnostic prescription	0.33		
Individualized instruction	0.32		
Individualized mathematics	0.32		
New science curricula	0.31		
Teacher expectations	0.28		
Computer-assisted instruction	0.24		
Sequenced lessons	0.24		
Advanced organizers	0.23		
New mathematics curricula	0.18		
Inquiry biology	0.16		
Homogeneous groups	0.10		
Programmed instruction	–0.03		
Class size	–0.09		
Mainstreaming	–0.12		

Note. The size of effects cannot be compared between the two publications. Effect sizes as reported by Fraser et al. (1987, p. 157) refer to differences between group means expressed in standard deviations. Effect sizes of Scheerens and Bosker (1997, p. 305) refer to log-transformed correlations (Fisher's *Z*) and should be interpreted according to Cohen's *d*. In addition, effect sizes refer to all kinds of student outcomes.

variables such as reinforcement, feedback, cooperative learning, differentiation/adaptive instruction, and time on task yielded the highest effect sizes.

Fraser et al. (1987) applied a rather detailed categorization that referred to specific instructional treatments (e.g., reading training, reading experiments, diagnostic prescription), domain-specific teaching approaches (e.g., science mastery, inquiry biology, new mathematics curricula), and more general teaching components (e.g., reinforcement, cues and feedback). Scheerens and Bosker's (1997) categorization was based on a set of time and mastery teaching models (Bloom, 1976; Caroll, 1963; Rosenshine, 1979). Thus, components such as time on task, structured teaching, opportunity to learn, and feedback were used to categorize teaching components.

When comparing both meta-analyses, two insights emerged. On one hand, some components of teaching (e.g., reinforcement, feedback, cooperative learning, mastery learning and learning time, adaptive instruction) turned out to be effective in both meta-analyses—that is, in meta-analyses that represented different decades of teaching effectiveness research. On the other hand, the two meta-analyses used different categorizations of variables based on research questions and underlying theoretical models. Consequently, these studies produced a diverse list of effective teaching components that restricted our ability to compare results. To simplify and provide continuity in teaching effectiveness research, we draw on Scheerens and Bosker's (1997) meta-analysis and use their categorization as an initial way to examine teaching effectiveness in the past decade (see their results in Section 6.1). We then examine teaching effectiveness using categorizations based on more recent models of teaching and learning.

Developments in Teaching Effectiveness Research in the Past Decade

Before outlining more recent models of teaching and learning, we note two important developments in the past decade that have influenced teaching effectiveness research. The first development is that researchers have concentrated on more global aspects of teaching and on analyzing teaching patterns or regimes instead of single teaching acts (Borko, 2004). Second, we found two distinct approaches to teaching effectiveness research of the last decade. In the first approach, researchers focused on large-scale surveys with sophisticated statistical models that controlled for extraneous variables, thereby increasing statistical power in detecting effects (Raudenbush, 2004; Rowan, Correnti, & Miller, 2002). Most of these studies were embedded in large-scale surveys that aimed to monitor student competencies and instructional practices (D. K. Cohen, Raudenbush, & Loewenberg Ball, 2003; Shavelson, Phillips, Towne, & Feuer, 1986).

In the second approach, researchers focused on processes of learning in specific knowledge domains (Brown, Collins, & Duguid, 1989; Collins et al., 1989; De Corte, Verschaffel, Entwistle, & Merrienboer, 2003; Greeno, Collins, & Resnick, 1996). This approach resulted in a growing number of quasi-experimental and experimental studies of the effects of specific instructional approaches on students' learning (D. K. Cohen et al., 2003). Moreover, these studies characterized student learning as multidimensional, meaning that instructional effects were examined on cognitive, affective, and metacognitive outcomes (Snow et al., 1980). Floden (2001) summarized this development as follows:

> When the work on teaching effects was under fire, research-supported recommendations for teaching were replaced by a variety of approaches, described variously (and loosely) as Socratic, progressive, constructivist, pragmatist, and Deweyian. These approaches all claim to be taking account of recent research on learning, which stresses the importance of students' active engagement. (p. 6)

From a cognitive perspective, then, teaching has been defined as the creation of learning environments in which students maximize the possibility of executing the cognitive activities necessary for building knowledge and reasoning capacity (Floden, 2001).

Cognitive Models of Teaching and Learning

In the past decade, models of teaching and learning have changed. Thus, we developed a second framework to account for these models in teaching effectiveness studies. The framework is based on three considerations: First, if the effectiveness of teaching is defined by the growth in student learning, our framework should be based on student learning. Second, the framework should allow us to categorize teaching variables according to underlying learning components. Third, the framework should be simple and functional yet categorize a broad variety of teaching variables into learning components. In what follows, we elaborate on these considerations: First, we summarize the state of the art on teaching and learning components. Second, we outline the framework of teaching and learning components that we used to classify research studies for this meta-analysis.

Components of Teaching and Learning

Most studies that investigated teaching effects on student learning in the past decade focused on intentional learning in an organized setting (e.g., school instruction). Furthermore, the components of learning referred to higher order learning: the development of meaning and in-depth understanding of learning content (Bransford et al., 2000). These studies, then, viewed learning in the following ways:

Constructive. Learning is a set of constructive processes in which the individual student (alone or socially) builds, activates, elaborates, and organizes knowledge structures. From this conception of learning, it follows that teaching should maximize the opportunity for students to engage in activities that promote higher order learning (Bransford et al., 2000; Brown et al., 1989; De Corte et al., 2003; Donovan & Bransford, 2005; Greeno et al., 1996).

Domain specific. Intentional learning (especially within a school context) is content specific and, thus, depends on the corresponding knowledge domain. Thus, teachers should create an environment in which students are able to engage in domain-specific learning activities. To do so, teachers need to have an in-depth understanding of the content and the nature of the domain that is pedagogically useful (pedagogical content knowledge; e.g., American Association for the Advancement of Science, 1993; Baumert, Blum, & Neubrand, 2004; Chi, Feltovich, & Glaser, 1981; National Research Council, 2001; OECD, 2004; Schoenfeld, 1985; Shulman, 1987; Wayne & Youngs, 2003; White & Frederiksen, 1998).

Social. The construction of knowledge takes place within the community of students in a classroom. In various classroom settings, students are encouraged to build knowledge within the community of learners, to explicate their knowledge, and to regulate and monitor their learning processes (Brown et al., 1989; Collins et al., 1989; Palincsar & Brown, 1984; Slavin, 1995).

Goal directed. Knowledge building in settings such as school instruction is intentional. Students either set individual goals or integrate externally given goals (e.g., teaching goals) as individual learning goals. In addition to goal setting, students have to focus on these learning goals to engage in goal-directed learning activities

(Ausubel, 1960; Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002; Rosenshine, 1979; Ryan & Deci, 2000).

Evaluative. Goal-directed, intentional learning has a component in which students and teachers evaluate achievement of corresponding learning goals. The evaluation of learning is manifold and implies self-evaluation and different kinds of formative and summative assessment (American Association for the Advancement of Science, 1990a, 1990b; Black & Williams, 2003; Donovan & Bransford, 2005; National Research Council, 1990, 1996, 2001).

Regulative. Learning includes a regulative component. Students should have the opportunity to internalize processes of stimulating, monitoring, and regulating learning. Teaching can, for example, guide and facilitate these regulative processes by providing scaffolding, feedback, and support or by teaching students strategies for self-regulation (Alexander, 2000; Artelt, Baumert, McElvany, & Peschar, 2003; Boekaerts, 1999; Kardash & Amlund, 1991; Pintrich & De Groot, 1990).

A Framework of Teaching and Learning Components

To integrate the components outlined above, we draw on a model developed by Bolhuis (2003). She makes distinctions between five interrelated components: goal setting, goal orienting, executing learning activities, evaluating, and regulating/monitoring/deciding. *Goal setting* and *goal orientation* refer to classroom activities that clarify goals and orient students toward learning to achieve the goals. Classroom learning activities attempt to impact students' internal (e.g., information-processing) and external (e.g., hands-on) activities. Evaluating involves those classroom activities that enable teachers and students to judge progress toward learning goals. Finally, the model includes a regulative component that is focused on those activities necessary to stimulate, monitor, and decide on learning. According to Bolhuis, the first four components are arranged in a cycle, whereas regulating/monitoring/deciding represents the central integrative component linking the other four. Furthermore, regardless of cyclical arrangement, the components do not necessarily play out in a particular sequence; paths can be characterized by slopes and moves between components. To build knowledge, however, students have to engage in learning activities. Bolhuis assumed that executing learning activities was the component most proximal to knowledge-building processes.

We adapted and expanded Bolhuis's (2003) model to use as a framework for our meta-analysis (Figure 1). We added four components: (1) knowledge domain, (2) amount of time for learning, (3) organizational frame for learning, and (4) classroom social climate. The remaining components of Bolhuis's model were (5) goal setting and orientation; (6) executing learning activities; (7) evaluation; and (8) regulation, monitoring, and decision-making. In addition, we distinguished three aspects of the executing learning activities component: (6a) social interactions/direct experiences, (6b) basic information processing, and (6c) domain-specific information processing. Social interactions and direct experiences include social, cooperative, and behavioral activities in a classroom. Basic information processing refers to the assumption that knowledge building involves cognitive activity to process verbal and other symbolic information. Domain-specific information processing refers to learning activities that are necessary and most adaptive for knowledge building in

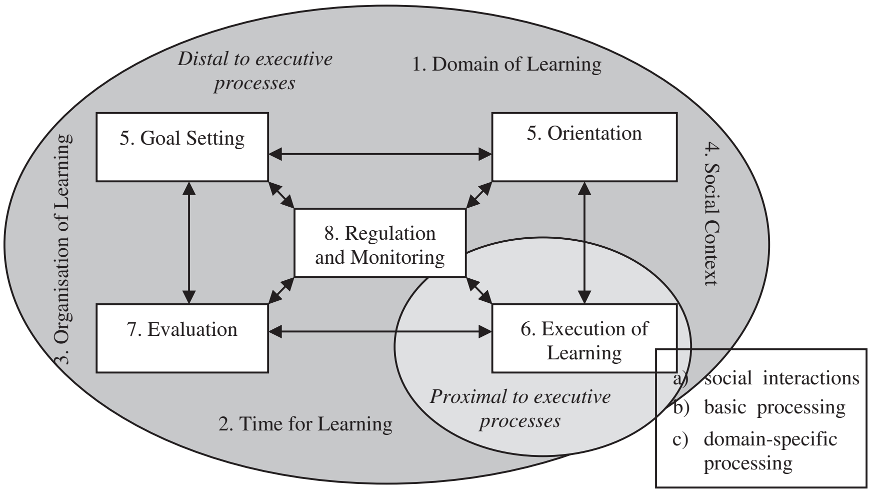

FIGURE 1. *Model of teaching and learning components.* Reprinted from *Learning and Instruction*, Volume 13(3), Sanneke Bolhuis, Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. Copyright (2003), with permission from Elsevier.

a domain. With additions, then, Bolhuis's model provided a framework in which we could map diverse teaching variables.

In the following, we illustrate these components from the perspective of teaching. These elaborations are based on theoretical considerations, as outlined in the previous section on components of teaching and learning and are further illustrated in the Method section of the meta-analysis.

From a teaching perspective, we consider four components that cover the instructional context and provide the necessary frame for students to engage in learning activities: (1) knowledge domain, (2) time for learning, (3) organization of learning, and (4) social context. The first component, knowledge domain, differentiates teaching and learning in different content areas (e.g., mathematics, reading, science). The second component, time for learning, takes into account the amount of time provided for teaching as a limiting frame for students to engage in learning activities. The third component, organization of learning, refers to the extent to which teachers provide an orderly and functional classroom setting (classroom management). And the fourth component, social context, focuses on the degree to which teachers establish a social learning climate within classrooms.

The central components of teaching that are conjectured to lead to student learning are (5) goal setting/orientation, (6) execution of learning activities, (7) evaluation, and (8) regulation and monitoring (Bolhuis, 2003). To the degree that these components are present in teaching, student learning is expected to increase. Teaching acts such as clarifying goals, teaching in a clear and structured way, or activating student pre-knowledge are important elements of the goal-setting and orientation component. The execution of learning activities is characterized by teaching acts that support social interactions between students and provide direct experiences for students, facilitate the basic processing of information (e.g., high language level, thinking-aloud methods), or provide domain-specific opportunities for processing content information (such as mathematic problem solving, science inquiry). Evaluation of learning characterizes teaching acts that aim to assess student progress toward learning goals. And the regulation and monitoring component includes teaching acts such as feedback and support or teaching students strategies of self-regulation and self-monitoring.

Conjectures Moving Forward in Reporting Our Meta-Analysis

Based on these theoretical considerations, we made the following conjectures for the meta-analysis:

Conjecture 1: Traditional models of teaching and learning. Previous meta-analyses have, appropriately, been based on models of teaching and learning dominant at that point in time. To provide continuity between our meta-analysis and these past meta-analyses, we should compare recent findings with past analyses using the older categories. We assume that studies of the past decade do not necessarily result in different effect sizes than those reported by previous meta-analyses. Moreover, given the theoretical developments in the past decade we have to assume that the set of investigated teaching variables has become broader and that traditional categorizations might not fully cover the broader range.

Conjecture 2: Recent models of teaching and learning. Based on the framework in Figure 1, our meta-analysis makes certain theoretical assumptions about the effectiveness of different teaching and learning components. We assumed that learning is a set of constructive processes in which the individual student (alone or socially) builds, activates, elaborates, and organizes knowledge structures. These processes are internal to the student and can be facilitated and fostered by components of teaching. Moreover, we assumed that higher order learning and a deep understanding of learning content is based on the quality of knowledge building and, thus, on the execution of learning activities. Learning activities should evoke both basic information processing and domain-specific processing. Consequently, we assumed the area of executing learning activities to be most proximal to knowledge building. Thus, we hypothesized that variables associated with the learning activities component would produce the greatest effects on student outcomes. We further assumed that teaching variables that cover the context of learning were more distal to knowledge building. They provide the necessary frame for students to engage in learning activities. Thus, we assumed that more distal variables would be associated with lower effect sizes compared to more proximal variables.

II. Methodological Approaches in Teaching Effectiveness Studies

We identified two distinct approaches to the study of teaching effectiveness in the past decade's research. The first approach, estimation of effects in large-scale surveys, addressed methodological challenges by increasing statistical power. The second approach, which stems from a cognitive psychological paradigm, used experimental and quasi-experimental intervention designs to investigate in-depth processes of knowledge building and its facilitation by teaching. In the following, these two approaches are reviewed from a methodological perspective to answer the question, "To what extent might effect sizes in meta-analyses be influenced by the choice of research design?"

Correlational Survey Designs

How might correlational survey designs influence the results of a meta-analysis on the effects of teaching on student learning? First, for a long time, researchers have debated about how teaching variables should be operationally defined in survey studies (for an overview, see Rowan et al., 2002). Because they tend to rely on proxy (self- or other-report) variables, we assumed that most of the teaching–learning components investigated in survey studies would be categorized as distal to executive learning activities (Figure 1).

Second, standard survey research has used statistical models in which a number of student prerequisite variables were taken into account before teaching effects were examined (value-added or covariate adjustment models). Those models usually result in low to moderate estimates of teaching-component effect sizes (D. K. Cohen et al., 2003; Raudenbush, 2004; Rowan et al., 2002). These relatively low effect sizes have been explained in two ways. First, proxy variables of questionable reliability and validity are used to measure teaching components. Second, variance decomposition models measured effects on status, not change in student outcomes:

> Instead, such analyses are simply modeling students' achievement status, which in a valued-added framework has been adjusted for students' social background and prior achievement. When viewed in this way, it is not surprising to find that teacher effects are relatively small in covariate adjustment models. Such models, in fact, are assessing teacher effects on achievement status, not change. (Rowan et al., 2002, p. 1530)

As a consequence, the authors argued, to properly estimate teaching effects, researchers need to estimate individual student growth curves. Rowan et al. (2002) have shown that effect sizes estimated by latent growth models were substantially higher than effect sizes estimated by covariate adjustment models (latent growth models: Cohen's $d = .72$ to .85; covariate adjustment models: $d = .21$ to .42). However, to date, the number of survey studies with a research design that allowed researchers to apply latent growth modeling is limited. Thus, we have to expect that the majority of studies will apply covariate adjustment models.

Summarizing the consequences of these considerations for our meta-analysis, we have to assume low effect sizes for correlational survey studies. The low effect sizes can arise due to distal measurement of teaching characteristics, the choice of covariate adjustment models, or simply the approach (survey) used to investigate natural variations in teaching practice. We cannot systematically disentangle the combination of these sources of low effect sizes within the scope of this meta-analysis. However, by reporting effect sizes from correlational surveys in the past decade, we might contribute to the discussion on the adequacy of research designs in teaching effectiveness research (D. K. Cohen et al., 2003; Raudenbush, 2005).

Quasi-Experimental and Experimental Intervention Studies

The second group of teaching effectiveness studies applied quasi-experimental and experimental research designs. Given the current discussion about the gold standard in American education research and the call for large-scale randomized field trials, it seemed appropriate to review quasi-experimental and experimental studies separately (Boruch & Mosteller, 2002; Campbell & Stanley, 1966; D. K. Cohen et al., 2003; National Research Council, 2004, 2005; Raudenbush, 2005). The number of experimental studies in the past decade, however, is rather limited (Cook & Payne, 2002; Crawford & Impara, 2001). Thus, investigating the effectiveness of experimental studies in educational research might be the task of a meta-analysis in the next decade but not primarily of the past decade. For this meta-analysis, we initially combined these two approaches and later compared the effect sizes of experimental and quasi-experimental studies. However, the main research question regarding the impact of research design on instructional effects focused on the comparison between correlational surveys and (quasi) experimental studies.

The objective and structure both of quasi-experimental and experimental studies are identical, except for the missing random assignment in quasi-experimental studies (Cook & Payne, 2002; Raudenbush, 2005; Shadish, Campbell, & Cook, 2002). Despite current knowledge about the design of quasi-experimental studies, we assumed that most of the studies of the past decade would not adequately address potential selectivity bias:

> Although it is true that the results of a well-designed quasi-experiment may not differ substantially from the results of a randomized experiment, the average quasi-experiment in education seems to be lamentable in the confidence it inspires in the causal conclusions drawn. Recent advances in the design and analysis of quasi-experiments are not getting into educational research, where they are sorely needed. (Cook & Payne, 2002, p. 173)

The most striking characteristic of (quasi) experimental research in the past decade might be its heterogeneity in all imaginable facets: The teaching variables studied varied from research on very specific training (e.g., metacognitive strategies) to evaluations of broad and intricate interventions (e.g., authentic pedagogy; Scheerens et al., 2005). The studies' theoretical backgrounds were often fuzzy, despite the fact that most of the studies referred to current cognitive models of teaching and learning. Indeed, each research group had given its teaching intervention a new name, resulting in a broad variety of terms for the system of teaching variables that were manipulated.

An additional problem is publication bias. The general interest in finding interventions that worked might have resulted in a publication bias toward positive effects of teaching interventions on student outcomes (Lipsey & Wilson, 1993). Lipsey and Wilson argue,

> Such research often yields an ambiguous mix of results—decidedly positive, suggestive, convincingly null, and hopelessly inconclusive. Research reviewers must then pick through these results with hopes of finding a preponderance of evidence supporting a conclusion about treatment efficacy. More specifically, they must attempt to sort and choose among studies on the basis of their methods, treatment variants, respondents, and the like to find those situations for which conclusions can be drawn. (p. 1181)

The advantage of (quasi) experimental studies is that, often, researchers dealt with the complexity of the field by concentrating on specific instructional treatments and attempted to provide valid measures of instructional characteristics and outcomes, often specifically tailored toward the goals of the teaching intervention (D. K. Cohen et al., 2003; Shavelson et al., 2003; Shavelson, Webb, & Burstein, 1986). With regard to the components of teaching and learning (see Figure 1), these studies were more likely to concentrate on measures that were proximal to both the teaching component and the executive learning processes. Given the choice and operationalization of teaching components and specific outcome measures, we might expect higher effect sizes than studies with distal measures. Thus, we intend to analyze correlational surveys and (quasi) experiments separately.

Conjectures About the Relationship Between Effect Size and Research Design

On the basis of the methodological considerations outlined above, we have made the following conjectures for the meta-analysis:

Conjecture: Correlational surveys. Correlational survey designs have been criticized for the limited validity of their measures of teaching components. Given the state of the art in survey studies, teaching components, although conceptually proximal, are actually measured distally to executive learning processes. Together with the predominant application of covariate adjustment models, we expect low effect sizes.

Conjecture: (Quasi) experimental studies. The field of (quasi) experimental studies is heterogeneous. Nevertheless, these studies are more specific in their object of investigation, choice of teaching component, and outcome measures than are surveys. Thus, we expect (quasi) experimental studies to be focused on teaching components that are more proximal than survey studies to executive learning activities and so to result in medium to high effect sizes.

III. Meta-Analysis of Teaching Effects on Student Learning

Research Questions

Based on the theoretical considerations outlined in Parts I and II, the following research questions were addressed in our meta-analysis:

General Teaching Effects on Student Learning

If teaching variables are classified according to previous meta-analyses, to what extent do current effect size estimates agree with those of a decade earlier?

What is the impact of teaching on student learning if teaching variables are classified on the basis of recent models of teaching and learning?

Do teaching and learning components more proximal to executive learning processes show higher effect sizes than more distal components?

Are the effect size estimates stable for the major domains of reading, mathematics, and science? Are they stable for elementary education, secondary education, or both?

Effect Sizes Estimated from Correlational Survey Designs

How are teaching variables distributed across components of teaching and learning in correlational survey studies? Do correlational designs cover components distal to executive learning processes?

How large are the effects of teaching on student learning as estimated in correlational survey studies (using Cohen's criterion for small, moderate, and large effect sizes)?

Do effect size estimates systematically vary with different measures of teaching components (teacher questionnaire, student questionnaire, observation or video analysis)?

Effect Sizes Estimated From (Quasi) Experimental Research Designs

How are teaching variables distributed across components of teaching and learning in experimental studies? Do (quasi) experimental studies cover components of learning that are domain specific and more proximal to executive learning processes?

How large are the effects of teaching on student learning as estimated from (quasi) experimental studies (using Cohen's criterion)?

Do experimental studies show greater effect sizes than do quasi-experimental studies?

Method

Literature Search Methods

The literature on teaching effectiveness was searched broadly, canvassing the Web of Science, ERIC (Education Resources Information Center), and ERA (Educational Research Abstracts) databases from the years 1995 to 2004. Furthermore, recent reviews of research and books on teaching effectiveness were searched. The searches employed the following keywords: *effective instruction, instructional effectiveness, direct instruction, teacher effectiveness, mastery learning, constructivist teaching, mathematics instruction, reading instruction, science instruction, classrooms, mathematics teaching, reading teaching, science teaching*. Each teaching keyword was crossed with each of the following output keywords: *achievement, competencies, interest, motivation, engagement, attainment.* Overall, 333 publications matched the crossed keyword combinations (teaching characteristics × student output). The publications represented a number of countries in which teaching effectiveness has been investigated.[1] Despite the fact that the review of the literature was exhaustive, we are aware that we still might have missed some relevant publications. However, we believe that the search was thorough enough to present a collection of publications that is representative of the field.

Inclusion Criteria

The meta-analysis focused on teaching variables found in the regular school system. Thus, studies were included that primarily investigated teaching from Grades 1 to 13.[2] Studies that specialized in kindergarten were excluded. The one exception to this rule was long-term investigations from kindergarten to higher levels of schooling. Furthermore, studies with a focus on teaching students with severe learning disabilities were not included. The number of publications was reduced to 276 after taking these criteria into consideration.

The 276 publications were then filtered using the criterion that the study must report empirical findings of teaching effects on student learning. Application of this criterion reduced the number of published articles to 125. Each of these 125 publications was examined as to the number of relationships between a teaching variable and a student outcome measure (replication). Thus, each relationship between a specific teaching component and a student learning outcome represented one replication for this teaching component (Cooper & Hedges, 1994). The number of replications was 1,665 and varied from 1 to 121 replications per study. We then applied a second criterion to the remaining (125) publications: Effects had to be adjusted for at least some kind of student prerequisite (e.g., pre-knowledge, aptitude, socioeconomic status). Application of this second criterion eliminated another 13 studies, resulting in the final number of 112 publications and 1,357 replications.

Classification

Classification of research studies according to previous meta-analyses. Of the 1,357 replications, we could classify 742 according to Scheerens and Bosker's (1997) meta-analysis (see Table 1). Given the description of teaching variables in the publications, however, we could not consistently differentiate between Scheerens and Bosker's three categories of time on task, opportunity to learn, and homework. In some studies, the amount of homework was either used as an indicator for opportunity to learn or as an indicator for time on task. Thus, we decided to summarize these three categories as time for learning. We could distinguish the other categories of reinforcement (such as tokens given to students), feedback (such as information to students about learning progress), cooperative learning, differentiation/adaptive instruction, and structured teaching. Overall, 178 replications were classified as learning time, 203 replications as structured teaching, 185 as cooperative learning, 67 as feedback, 55 as reinforcement, and 54 as differentiation/adaptive instruction. We could not classify 598 replications into Scheerens and Bosker's categories; they were excluded from this first analysis. These excluded replications referred to teaching variables such as self-regulation, domain-specific teaching activities, teacher expectations, learning climate, classroom management, or evaluation of learning.

Classification of research studies according to the components of teaching and learning. Of the 1,357 replications, we classified 1,352 according to their predominant underlying component of teaching and learning (see Figure 1). Five replications could not be classified and were excluded from the analysis. These replications referred to a study using an integrative teaching approach in which a number of different teaching variables were applied in a combined research design, and these could not be disentangled according to the underlying teaching and learning components. In Table 2, we provide an overview of teaching and learning components and the teaching variables included in each component for the 1,352 replications.

The first component, time for learning, included studies that focused on time used for learning tasks (studies with keywords such as *time on task* or *opportunity to learn*) or time used for homework. Overall, 34 publications with 178 replications were added to the database.

The second component, organization of learning, comprised 17 publications with a focus on an orderly and functional classroom setting. Thus, 121 replications investigating instructional characteristics such as classroom management or discipline/control were summarized in this component.

TABLE 2
Integration of teaching variables into teaching and learning components

Teaching and learning component	Teaching variables	Number of publications	Number of replications	Number of replications with student background adjustment	Percentage of studies with student background adjustment
Time for learning	Time on task	34	196	178	91
	Opportunity to learn				
	Homework				
Organization of learning	Classroom management	17	143	121	85
	Discipline/control				
Social context for learning	Learning climate	20	136	113	83
	Learning orientation				
	High expectations				
	Teacher beliefs				
Goal-setting/orientation	Goal directed	33	224	155	69
	Clear and structured				
	Activating student prerequisites				
	Anchors/contexts				
Execution of learning activities					
Social interaction/ direct experience	Various teaching methods	33	222	202	91
	Cooperative learning				
	Student discussion				
	Student work, hands-on				

Basic information processing	Cognitive activation Active student engagement High language level Thinking aloud training	29	270	213	79
Domain-specific information processing	Math problem solving Scientific inquiry Reading/writing strategies	18	129	112	87
Evaluation of learning	Assessments/tests	10	88	87	99
Regulation and monitoring	Support Feedback and monitoring Adaptivity Self-regulation	32	257	171	67
Total		112	1,665	1,352	81%

Note. Due to the fact that some publications provided data for more than one category, the sum of the number of publications in the column does not correspond to the total number of publications investigated (112).

The third component, social context for learning, focused on the social learning climate in the classroom and teachers' expectations and beliefs about their students' learning. Twenty publications and 113 replications addressed this component.

The fourth component, goal setting and goal orientation, investigated clarity of goals, clear and structured teaching, activation of student pre-knowledge, or use of anchors and contexts in explaining learning contents. We found such variables in a total of 33 publications with 155 replications.

The fifth component, execution of learning activities, was divided into three subcategories: (a) social interactions/direct experiences, (b) basic processing of information, and (c) domain-specific information processing. For social interactions/direct experiences, publications with a focus on cooperative learning, student hands-on activities, student discussions, or the use of a variety of teaching methods were clustered. Thirty-three publications with 202 replications matched the description.

The component basic information processing was found in 29 studies (213 replications). These studies focused on providing opportunities to cognitively engage students in the learning content. Thus, we considered key variables such as cognitive activation, active cognitive student engagement, use of a high language level to engage students in higher order thinking, or think-aloud training with the goal of activating students' thinking to be basic information processing. Our purpose in summarizing the publications in this area was to focus on teaching variables aimed at stimulating students' in-depth elaboration and organization of learning content. These studies could be distinguished from those examining domain-specific processing in that they focused on generalizable factors in the processing of learning content.

Publications with a focus on domain-specific components in processing information were categorized as domain-specific processing. This component included 18 publications and 112 replications that investigated activities such as mathematical problem solving, scientific inquiry, or specific reading and writing strategies.

The sixth component, evaluation of learning, summarized variables from 10 publications (87 replications) with a focus on the evaluative components of learning. In the end, these exclusively investigated the effect of assessment on student outcomes. Thus, studies narrowly conceptualized evaluation and so restricted our inferences about the evaluation component.

Finally, the seventh component refers to 32 publications (171 replications) with a focus on the regulation and monitoring of student learning. The instructional variables clustered here included teacher support, feedback, monitoring, teaching adaptivity, and support of students' self-regulation of learning.

Classification of teaching effects by type of student outcome. Teaching effects on student learning were further differentiated according to three outcome measures: learning processes, motivational–affective, and cognitive. Outcome measures were categorized as learning process if they focused 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). Motivational–affective and cognitive outcomes referred to the long-term results of learning. Motivational–affective outcomes comprised motivational, affective, and conative results (e.g., development of stable interests, motivational orientations, attitudes or belief systems). Cognitive outcomes referred to results of learning with respect to the development of knowledge, measured either by standardized achievement and competency tests or specific tests of content understanding or student performance.

Classification by methodological characteristics. The replications were further categorized according to the methodological characteristics of the study. The following categories were employed:

Design: correlational, quasi-experimental, and experimental.
Measurement of instructional variables: (quasi) experimental manipulation (and implementation checked), teacher questionnaire, student questionnaire, principal questionnaire, and observation or video analysis.
Stage of schooling: elementary or secondary (middle and high).
Domain: reading, mathematics, or science.

Meta-Analysis Procedures

The meta-analysis was carried out by computing Fisher's Z as an index of effect size (Rosenthal, 1994), an index that is interpreted like a conventional correlation coefficient (r). All statistics reported in the original research articles (correlations, regression coefficients, t values, etc.), then, were transformed into this index.[3] Furthermore, the effect sizes were weighted by sample size (factor w = inverse of variance; Hedges & Olkin, 1985).[4] We only report effect sizes based on five or more replications.

The best known effect size coefficient, Cohen's d, is related to the correlation coefficient (and, hence, Fisher's Z) in the following way: $d = 2r/\sqrt{(1 - r^2)}$. In addition, a close estimate is $d = 2r$. To interpret the size of effects, J. Cohen (1969) considered a value of .20 as a small effect, a value of .50 as moderate, and a value of .80 as large. The fact that effect sizes may appear small, however, does not necessarily mean that they are unimportant. This is especially the case in the study of teaching and learning, given the large number of factors affecting students' performance in school systems. Moreover, even a small effect has consequences for thousands of students. Thus, the size of effects should be interpreted carefully. Moreover, the research questions driving this meta-analysis were less about the size of effects and more about underlying influences giving rise to the overall effect sizes. Consequently, we have no reason to believe that studies of the past decade should result in different effect sizes than those reported by previous meta-analyses where variables overlap (Fraser et al., 1987; Scheerens & Bosker, 1997).

Results

General Teaching Effects on Student Learning

Effects of teaching variables: Process–product paradigm meta-analyses. Our first research question addressed the impact of teaching on student learning if teaching components were classified according to Scheerens and Bosker's (1997) previous meta-analysis. All replications that we could classify into Scheerens and Bosker's model were used in estimating weighted effect sizes. In Table 3, the mean effect

TABLE 3
Mean effect sizes for teaching components as categorized by Scheerens and Bosker (1997): All outcomes and learning processes, motivational–affective, and cognitive outcomes

	All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	178	.04	8	.14	13	.12	157	.03
Structured teaching	203	.02	29	.04	39	.07	135	.02
Cooperative learning	185	.01	12	.11	24	.13	149	.00
Feedback	67	.01			16	.11	51	–.01
Reinforcement	55	.02	17	.05	14	.07	24	.01
Differentiation/adaptive instruction	54	.04			6	.04	44	.03

Note. *n* = number of replications; ES = weighted effect size on the basis of Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section. Scheerens and Bosker's categories of time on task, opportunity to learn, and homework have been summarized as time for learning (see their Section 5.3). The mean effect size of the three categories in Scheerens and Bosker's meta-analysis on all student outcomes is .11.

sizes are given for all student outcomes and are differentiated according to outcome: learning processes, motivational–affective, and cognitive.

Consistent, small effects (according to Cohen's *d* criterion) were found for each teaching variable (ranging from .01 to .04). The highest effect sizes were found for time for learning (.04) and differentiation/adaptive instruction (.04) followed by structured teaching (.02), reinforcement (.02), cooperative learning (.01), and feedback (.01). Given these results, we consequently focused on factors giving rise to these low effect sizes.

One possible factor was aggregation: By aggregating effect sizes over very different teaching variables and student outcomes, we might have masked more specific effects. For example, effect size magnitudes might change if different student outcomes were considered. As shown in Table 3, this was the case. For the learning processes outcome, the effect size for time for learning was .14 followed by cooperative learning (.11). With respect to motivational–affective outcomes, the effect sizes were .13 for cooperative learning, .12 for time for learning, and .11 for feedback. Finally, the impact on cognitive outcomes was .03 for time for learning and for differentiation/adaptive instruction and .02 for structured teaching. Thus, overall low effect sizes were due to low effects of teaching on cognitive outcomes.

Effects of teaching variables: Cognitive models meta-analyses. The second research question addressed the impact of teaching on student learning if teaching variables were classified according to recent cognitive models of teaching and learning—teaching and learning components models. We expected that components proximal to executive learning processes would show higher effect sizes than more distal components would (see Figure 1). In Table 4, the mean effect sizes are given for each student outcome—learning processes, motivational–affective, and cognitive.

TABLE 4
Mean effect size for teaching variables categorized according to teaching and learning components: All outcomes and learning processes, motivational–affective, and cognitive outcomes

	All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	178	.04	8	.14	13	.12	157	.03
Organization of learning	121	.01	9	.01	26	.06	86	.00
Social context	113	.04	6	-.03	35	.01	72	.05
Goal setting and orientation	155	.03	38	.09	19	.07	98	.02
Execution of learning activities								
Social/direct experiences	202	.01	21	.11	24	.13	157	.00
Basic processing	213	.02	21	.05	41	.08	151	.01
Domain-specific processing	112	.21	19	.16	15	.21	78	.22
Evaluation of learning	87	.01			15	.00	72	.02
Regulation/monitoring	171	.02	17	.05	40	.08	114	.01

Note. *n* = number of replications; ES = weighted effect size on the basis of Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section.

Prior studies have shown that the highest effect sizes were reported for variables proximal to executive learning activities. Replications with a focus on domain-specific learning activities have resulted in the highest effect size on all outcomes (.21). For cognitive, motivational–affective, and learning process outcomes, the highest effect sizes were found for the execution of domain-specific activities (.22, .21, and .16, respectively). On one hand, these findings indicated that domain-specific activities had the highest teaching effects on student outcomes. On the other hand, the findings also showed that teaching effectiveness was diverse and that the impact of teaching depended on the outcome measured. Some teaching components affected motivational–affective outcomes, others learning processes, or still others cognitive outcomes.

Effect sizes in reading, mathematics, and science. Effect sizes in reading, mathematics, and science are reported in Table 5. Domain-specific learning activities, once again, produced the largest effects on cognitive learning outcomes in these three academic domains: .15 for reading, .18 for mathematics, and .30 for science. Moreover, the pattern of effect sizes was similar across the three domains.

Effect sizes in elementary and secondary education. The effect sizes in elementary and secondary education are reported in Table 6. Similar patterns emerged for elementary and secondary education. Domain-specific learning activities produced the highest effect sizes both for elementary (.20) and secondary education (.21). In elementary education, teaching components such as time for learning (.11), organization of learning (.08), social experiences (.05), and domain-specific activities (.20) produced the highest effect sizes. In secondary education, the effect sizes

TABLE 5
Mean effect size by outcome: All outcomes and effects on learning processes, motivational–affective, and cognitive outcomes

	Reading								Mathematics								Science							
	All outcomes		Learning processes		Motivational–affective		Cognitive		All outcomes		Learning processes		Motivational–affective		Cognitive		All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	60	.19					54	.21	88	.04			6	.18	82	.03	10	.01	6	.15				
Organization of learning	5	.01					5	.01	69	.02			21	.06	48	.01	16	–.01					8	-.01
Social context	47	.06			9	.05	38	.07	45	.03			14	.09	31	.03								
Goal setting and orientation	23	.03					22	.04	83	.03	6	–.03	13	.08	64	.02	30	.06	21	.06			6	.11
Execution of learning activities																								
Social/direct experiences	82	.03					80	.03	51	-.02			9	.01	42	-.02	62	.01	13	.10	14	.20	35	.00
Basic processing	33	.01					33	.01	144	.02	7	.03	35	.08	102	.01	23	.03					15	.03
Domain-specific processing	39	.16	7	.10			28	.15	35	.19	10	.20			22	.18	35	.26			7	.17	28	.30
Evaluation of learning	26	.02					26	.02	38	.03					34	.03								
Regulation/monitoring	48	.04			6	.07	42	.04	72	.01			22	.10	50	-.04	15	.03	15	.03				

Note. *n* = Number of replications; ES = weighted effect size on the basis of Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section.

TABLE 6

Mean effect sizes across stages of schooling: All outcomes and learning processes, motivational–affective, and cognitive outcomes

	Elementary								Secondary							
	All outcomes		Learning processes		Motivational–affective		Cognitive		All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	124	.11			8	.11	114	.12	54	.01	6	.15	5	.24	43	.01
Organization of learning	6	.08					5	.10	115	.01	9	.01	25	.06	81	.00
Social context	83	.04	6	-.03	22	.01	55	.05	30	.03			13	.04	17	.03
Goal setting and orientation	69	.03					63	.03	86	.04	35	.09	16	.08	35	.02
Execution of learning activities																
Social/direct experiences	101	.05					96	.05	101	.00	20	.12	20	.14	61	–.01
Basic processing	77	.03					74	.03	136	.02	19	.06	40	.08	77	.00
Domain-specific processing	49	.20	9	.13	8	.30	32	.20	63	.21	10	.20	7	.17	46	.23
Evaluation of learning	52	.02			11	.00	41	.02	35	.01					31	.01
Regulation/monitoring	77	.04			16	.06	60	.03	94	.01	16	.05	24	.09	54	–.01

Note. n = number of replications; ES = weighted effect size on the basis of Fisher’s Z. To interpret effect sizes according to Cohen’s d, use the formula presented in the Method section.

were generally smaller for most teaching components (goal setting and orientation, .04; social context, .03; basic processing, .02) than they were in elementary education.

Summary

The analysis of teaching effects over the past decade, based on a categorization of variables from the process–product paradigm (Scheerens & Bosker, 1997), yielded generally small effect sizes (Cohen's $d = .04$) with little variation between the effects of different teaching components ($d = .02$ to .08). The first impression in interpreting these findings (compare Table 1) might be that teaching effectiveness studies from the past decade produced smaller teaching on student learning than did findings from studies in previous decades. This conclusion, however, is misleading because we lack systematic information about differences in the statistical procedures (e.g., the weighting for sample sizes), inclusion criteria (e.g., the control of student background characteristics), literature search methods, and study composition (e.g., surveys, experiments). Thus, we turned our attention to identifying the influences giving rise to these low effect sizes. The paramount influence turned out to be aggregation—by disaggregating effects by type of outcome, we found the low effects to be due to cognitive outcomes.

We expected that by categorizing teaching components on the basis of recent cognitive models of teaching and learning, we would find higher effect sizes. Consistent with this expectation, we found that providing opportunities to execute domain-specific learning activities yielded moderate effect sizes (all outcomes: Fisher's $Z = .21$, $d = .41$), regardless of domain (science, mathematics, reading), stage of schooling (elementary or secondary), or type of outcome (learning processes, cognitive, motivational–affective). This finding is in line with the theoretical conjecture that teaching conditions in which students are able to execute domain-specific learning activities should have a large effect on cognitive aspects of learning. We found, however, that executing domain-specific activities was also one of the most important factors for motivational–affective outcomes (e.g., interest or self-concept of ability). We predicted that other more distal components such as the organization of learning would result in smaller effects on learning outcomes. Again, the meta-analysis provided support for this conjecture.

Overall, the results so far shed light on the general distribution of effects across teaching and learning components. However, at the same time, methodological issues arise. What if studies that investigated domain-specific learning activities were more likely to be related to a specific research design? What if studies of domain-specific activities had more valid and reliable measures of instructional characteristics than did other components? We now turn our attention to these questions. First, we summarize findings from correlational survey designs, then focus on (quasi) experiments, and conclude by comparing effect sizes from experimental and quasi experimental designs.

Teaching Effects on Student Learning by Research Design

Correlational Survey Designs

We addressed three questions with regard to correlational survey designs: (a) How were teaching variables distributed across components of teaching and learning,

TABLE 7
Mean effect sizes for correlational surveys: All outcomes and learning processes, motivational–affective, and cognitive outcomes

	Correlational surveys							
	All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	142	.05			8	.13	133	.04
Organization of learning	102	.01			26	.06	72	.00
Social context	92	.04	6	–.03	31	.01	55	.05
Goal setting and orientation	97	.03	15	.12	15	.08	67	.01
Execution of learning activities								
Social/direct experiences	79	.00			17	.14	58	–.01
Basic processing	162	.02	16	.04	38	.09	108	.01
Domain-specific processing	4							
Evaluation of learning	84	.01			15	.00	69	.02
Regulation/monitoring	143	.02	5	.13	34	.08	104	.01

Note. *n* = number of replications; ES = weighted effect size on the basis of Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section.

and do correlational designs cover more distal than proximal executive learning processes? (b) How large were the effects of teaching on student learning (following Cohen's criterion for small, moderate, and large effect sizes)? (c) Do effect size estimates systematically vary with different types of teaching component measures (teacher questionnaire, student questionnaire, observation or video analysis)?

Distribution of correlational survey replications. The distribution of replications across learning components is given in Table 7. Of the replications, 67% came from correlational surveys ($N = 905$). The majority of correlational surveys measured distal components: 660 such replications (sum of time for learning, organization of learning, social context, goal setting, evaluation of learning, and regulation) compared to 245 replications for the executing learning activities component (and only 4 replications were in domain-specific processing). Thus, the general effects reported in the first section might be misleading. As a consequence, effect sizes should be disaggregated by research design.

Effects of correlational surveys on student outcomes. The effect sizes reported in Table 7 are consistent with those reported in the preceding section of this article. In fact, the pattern of effect sizes across teaching and learning components, as well as across different outcome measures, was comparable with the generally low effect sizes.

Operationalization of instructional variables. Finally, we further differentiated within correlational survey studies as to the source of the teaching component measurement (e.g., questionnaire, video). The question is, Do effect sizes differ

TABLE 8
Correlational surveys: Effect sizes of different measures for instructional variables on all student outcomes

	Measurement of instructional variables					
	Teacher questionnaire		Student questionnaire		Observation/ video analysis	
	n	ES	*n*	ES	*n*	ES
Time for learning	33	.00	19	.12	9	.01
Organization of learning	13	.01	84	.01		
Social context	19	.08	59	.04	7	.15
Goal setting and orientation	16	.02	33	.06	26	.00
Execution of learning activities						
Social/direct experiences			57	-.01	15	.02
Basic processing			121	.02	26	.04
Domain-specific processing						
Evaluation of learning	26	.01	41	.02		
Regulation/monitoring	22	.00	48	.08	10	.28

Note. *n* = number of replications; ES = weighted effect size on the basis of Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section.

as to how teaching components were measured? To address this question, each replication was coded according to its source of measurement: teacher questionnaire, student questionnaire, school principal questionnaire, or classroom observation/ video analysis. Table 8 reports the corresponding effect sizes. The source, school principal questionnaire, did not occur and, thus, was not included in the table.

Observational and video analysis measures produced the highest effect sizes: regulation and monitoring (.28) followed by the social context of learning (.15). The next highest effect sizes were found with the student questionnaire: .12 for time for learning, .06 for goal-setting and orientation, and .08 for regulation/ monitoring. Finally, the highest effect sizes based on teacher questionnaires was found for the social context of learning: .08 (including variables such as teacher expectancies and beliefs).

Summary

Correlational surveys tended to focus on distal teaching and learning components, as indexed by number of replications of variables. Correlational studies resulted in the same low effect sizes reported in preceding sections. Furthermore, correlational surveys did not provide a particular advantage in investigating specific teaching and learning components. In fact, the effect sizes were low for each of the components. Finally, we found that effect size magnitudes varied systematically with the data source (e.g., observation, teacher questionnaire, or student questionnaire). As a consequence, we suspect that a mix of data and not a single method might be especially fruitful for future survey studies.

TABLE 9
Mean effect sizes for (quasi) experiments: All outcomes and learning processes, motivational–affective, and cognitive outcomes

	(Quasi) experiments							
	All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	36	.03	7	.15	5	.05	24	.01
Organization of learning	19	.28	5	.10			14	.33
Social context	21	–.01					17	.00
Goal setting and orientation	58	.10	23	.05			31	.16
Execution of learning activities								
Social/direct experiences	123	.06	17	.13	7	.08	99	.05
Basic processing	51	.03	5	.09			43	.02
Domain-specific processing	108	.23	19	.16	15	.21	74	.25
Evaluation of learning	3							
Regulation/monitoring	28	.05	12	.03	6	.11	10	.03

Note. *n* = number of replications; ES = weighted effect size on the basis of Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section.

Quasi-Experimental and Experimental Research Designs

Finally, we addressed the impact of teaching and learning components on student outcomes in quasi-experimental and experimental studies. We first report the distribution of (quasi) experimental studies across teaching and learning components. Our hypothesis was that teaching components proximal to executive learning processes would be investigated more frequently in (quasi) experimental designs than in correlational designs. This conjecture was supported above, but we provide more detail here. We then report effect size findings, conjecturing that (quasi) experimental designs would result in moderate effect sizes. This conjecture was based on the fact that proximal components were investigated more frequently in (quasi) experimental designs than in correlational designs. Finally, we compare effect sizes estimated in experimental and quasi-experimental studies.

Distribution of (quasi) experimental replications. The distribution of (quasi) experimental replications across learning components is reported in Table 9. Of the replications in our database, 33% came from quasi-experimental or experimental designs ($N = 447$). We found, as expected, that 63% of the replications referred to execution of learning activities and, thus, to proximal components. The majority of replications within this component investigated the role of either social experiences (44%) or domain-specific activities (38%).

Effects of (quasi) experiments on student outcomes. Effect sizes ranged from –.01 for social context to .28 for organization of learning (Table 9). Domain-specific activities (.16), time for learning (.15), social experiences (.13), and the organization of learning (.10) had the highest impact on the learning processes outcome.

Domain-specific activities (.21) and regulation/monitoring (.11) showed the highest effect sizes for motivational–affective outcomes. Cognitive outcomes were most strongly influenced by domain-specific learning activities (.25), organization of learning (.33), and goal setting and orientation (0.16). Overall, effects sizes of (quasi) experimental studies were larger than those of correlational surveys (range of effect sizes for correlational surveys: .00 to .05; see Table 7).

Effects sizes of quasi-experiments and experiments. Effect sizes are disaggregated for quasi-experiments and experiments in Table 10. Overall, the pattern of effect sizes across teaching and learning components was similar for quasi-experimental and experimental studies, with the latter yielding considerably higher estimates, as predicted. In both cases, domain-specific learning activities showed the greatest effects on student outcomes. The effect sizes produced by experiments were generally larger than those of quasi-experiments (all outcomes: quasi-experiments = .19, experiments = .27; learning processes: quasi-experiments = .15, experiments = .40; cognitive: quasi-experiments = .20, experiments = .30). With regard to motivational–affective outcomes, quasi-experiments yielded larger effect sizes than did experiments (motivational–affective: quasi-experiments = .28, experiments = .18). However, motivational–affective outcomes were few in number, and the results have to be interpreted carefully.

Although (quasi) experimental studies produced higher effect sizes than did correlational surveys, on average, they did not incorporate a number of important teaching components. For quasi-experimental studies, low numbers of replications were found for the following components: organization of learning, social context of learning, and evaluation of learning. Experimental studies lacked measures of time for learning, evaluation of learning, and regulation and monitoring. Furthermore, motivational–affective outcomes were hardly investigated at all in (quasi) experimental studies. Thus, in the future, quasi-experimental and experimental studies might examine a more comprehensive set of teaching components relevant to the complete cycle of learning.

Summary

(Quasi) experimental studies resulted in small to medium effect sizes. In addition, the mean effect sizes of experimental studies were larger than for quasi-experimental studies. Providing students with opportunities to execute domain-specific learning activities yielded the highest effect sizes, regardless of quasi-experimental or experimental designs and regardless of the type of student outcome (learning processes, motivational–affective, cognitive). Both quasi-experimental and experimental studies focused on cognitive outcomes. Motivational–affective outcomes and learning processes were investigated less intensively. Finally, most replications of quasi-experimental and experimental designs referred to the executing learning activities component. Thus, the set of teaching components relevant to the complete cycle of learning was restricted in (quasi) experimental studies.

IV. Conclusion

What are the main findings of our meta-analysis of studies of the effects of teaching on student learning over the past decade? First, our teaching and learning

TABLE 10

Mean effect sizes for (quasi) experiments: All outcomes and learning processes, motivational–affective, and cognitive outcomes

	Quasi-experiments								Experiments							
	All outcomes		Learning processes		Motivational–affective		Cognitive		All outcomes		Learning processes		Motivational–affective		Cognitive	
	n	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES	*n*	ES
Time for learning	33	.05	7	.15	5	.05	21	.03								
Organization of learning									18	.28	5	.10			13	.33
Social context									17	.00					17	.00
Goal setting and orientation	37	.09	21	.06			13	.17	21	.12					18	.15
Execution of learning activities																
Social/direct experiences	31	.15	9	.13			19	.15	92	.04	8	.14			80	.03
Basic processing	19	.03					16	.05	32	.03					27	.02
Domain-specific processing	59	.19	14	.15	5	.28	40	.20	49	.27	5	.40	10	.18	34	.30
Evaluation of learning																
Regulation/monitoring	28	.05	12	.03	6	.11	10	.03								

Note. n = number of replications; ES = weighted effect size on the basis of Fisher's Z. To interpret effect sizes according to Cohen's d, use the formula presented in the Method section.

model, based on recent cognitive models of teaching and learning, proved useful in analyzing studies and effect sizes. We were better able to account for variations in effect sizes than were past models. Second, we found that the effects of teaching on student learning were diverse and complex but fairly systematic. Third, and more specifically, we found the largest teaching effects for domain-specific components of teaching—teaching components most proximal to executive learning processes. Fourth, by taking into account research design, we were able to disentangle meta-analytic findings. For example, we found that domain-specific teaching components, those components associated with the largest effect sizes, were mainly studied with quasi-experimental or experimental designs. Fifth, teaching effectiveness studies in the past decade were quantitatively predominated by correlational survey studies. These studies resulted in overall low effect sizes. In the following, we elaborate on these conclusions.

Models Drive Outcomes of Meta-Analyses on Teaching Effectiveness

We examined current teaching effectiveness research using models of teaching and learning as lenses. Two frameworks were applied: The first framework was built on a process–product model that focused on time and mastery and on teaching acts such as reinforcement and teacher support (Bloom, 1976; Caroll, 1963; Gage, 1963; Rosenshine, 1979). The second framework drew from recent cognitive componential models of teaching and learning (Bolhuis, 2003).

The process–product model has been widely used in meta-analyses and reviews of teaching effectiveness (Brophy, 2000; Brophy & Good, 1986; Creemers, 1994; Gage, 1963; Scheerens, 2000; Scheerens & Bosker, 1997; Scheerens et al., 2005; Walberg, 1986; Walberg & Paik, 2000). Scheerens and Bosker's study, the most recent major meta-analysis in this tradition, used variables such as time for learning, structured teaching, cooperative learning, feedback, reinforcement, and differentiation/adaptive instruction. They reported an overall mean effect size for teaching (p. 305) of .25, with effect sizes ranging from small (.06) to large (.58). We applied the process–product model to our data set of studies conducted during the past decade. In contrast to previous findings, we found a mean effect size of .02, with little variation across teaching components (.01 to .04).

Our first impression in seeing these findings was that studies of the past decade have resulted in a smaller impact of teaching on student learning than reported by studies from previous decades. This conclusion, however, is misleading. Despite the fact that we used the same theoretical model as Scheerens and Bosker (1997), and that we used the same index to report effect sizes (Fisher's Z), the two meta-analyses are not directly comparable. Our meta-analysis differed from those in the past because we weighted effect sizes for sample size and included only studies with controls for student prerequisites. In addition, our literature search might have differed from those in the past, but we lacked information to confirm this conjecture. Thus, in comparing and interpreting the findings of the two meta-analyses, readers have to be very careful.

Consequently, we focused our meta-analysis on ferreting out the underlying influences giving rise to our results. We began by creating a framework (Figure 1) consistent with recent cognitive componential models of teaching and learning (Bolhuis, 2003; Bransford et al., 2000; Brown et al., 1989; Collins et al., 1989; De

Corte et al., 2003; Donovan & Bransford, 2005; Greeno et al., 1996). We used that framework to categorize teaching variables into components to see if these components might explain effect size differences. We conjectured that components most proximal to knowledge-building executive processes should result in higher effects on student learning than do components distal to these processes. By applying this second model, we found that our meta-analysis produced a broader range of teaching effect sizes than had emerged when the first model was used (range = .01 to .21). Thus, our findings on differences between the two models showed that models drive the outcomes of meta-analyses. We stress the importance of making underlying theoretical models explicit in meta-analysis research.

Teaching Effectiveness Is Complex

In our meta-analysis, we recognized that the effects of teaching on student outcomes were diverse. We thought that some teaching components might have an effect on the students' cognitive growth, others on the students' motivational development; some might show short-term effects on learning processes, and others might show long-term effects (Snow, Corno, & Jackson, 1996; Snow et al., 1980). Thus, the meta-analysis investigated effects of teaching on three outcome measures: learning processes, cognitive outcomes, and motivational–affective outcomes. The majority of studies in the past decade, however, concentrated on cognitive outcomes.

We found that the execution of domain-specific learning activities had the strongest impact on cognitive outcomes, with a moderate effect size: $d = 0.45$ (J. Cohen, 1969).[5] For motivational–affective outcomes, highest effect sizes were associated with the following components: domain-specific activities, social experiences, time for learning, and regulation and monitoring. For learning processes, domain-specific learning activities, time for learning, and social experiences showed the highest effect sizes. Domain-specific activities consistently represented the most important influence of teaching on student learning and stood out from the other components.

Domain-Specific Components as the Most Effective Teaching Factor

Providing opportunities for students to engage in domain-specific learning activities was shown to be the component with the highest effect sizes, regardless of domain (reading, mathematics, science), stage of schooling (elementary, secondary), or type of learning outcome (learning processes, motivational–affective, cognitive). This finding is in line with the cognitive componential models that we used: Teaching conditions in which students are able to execute domain-specific learning activities are most likely to affect cognitive aspects of learning (De Corte, Greer, & Verschaffel, 1996; Mevarech & Kramarski, 1997; Reusser & Stebler, 1997; White & Frederiksen, 1998). We also found that executing domain-specific activities played an important role for motivational–affective outcomes (e.g., interest or self-concept of ability). However, the number of studies that included motivational–affective outcomes was considerably smaller than those investigating effects on cognitive outcomes. Researchers might consider investigating the effects of domain-specific teaching on learning processes and motivational–affective outcomes in more depth than is currently practiced.

Domain-Specific Teaching Studied Primarily Under (Quasi) Experimental Conditions

By disentangling overall findings by underlying research design, we found that domain-specific teaching components were mainly studied under quasi-experimental or experimental conditions. For our meta-analysis, we examined these two approaches jointly and then tentatively compared the effect sizes of experimental and quasi-experimental studies, recognizing limitations due to the small number of studies and replications. Our findings indicated that experimental studies produced higher effect sizes, in general, than did quasi-experimental studies. For many of the postulated teaching and learning components, we found that replications of (quasi) experimental studies were few or nonexistent, resulting in a number of components for which no effect sizes could be reported. Both quasi-experimental and experimental studies focused on investigating the teaching component: providing opportunities to execute learning activities. Thus, in the future, quasi-experimental and experimental studies might examine a more comprehensive set of teaching components relevant to the complete cycle of learning (Figure 1).

Majority of Teaching Effectiveness Studies Based on Correlational Survey Studies

Teaching effectiveness studies of the past decade, however, were numerically dominated by correlational survey studies as compared to (quasi) experimental studies. Two thirds of the investigated relationships between a teaching component and a student outcome in the meta-analysis were based on correlational surveys. The potential of correlational survey studies for explaining teaching effects on student learning have been questioned, with criticisms pointing to the use of distal measures of teaching components (Crawford & Impara, 2001; Shadish et al., 2002; Shavelson et al., 1986), the predominance of covariate adjustment models (Raudenbush, 2004; Rowan et al., 2002), and the approach taken in interpreting natural variations in teaching (D. K. Cohen et al., 2003; Rowan et al., 2002). (We could not disentangle these criticisms to determine which best explained our findings.) The findings supported our prediction of small effect sizes for correlational survey studies. We found that effect size magnitudes varied with the data source (e.g., teacher questionnaire, student questionnaire, observation or video analysis); however, the predominant data source for measuring teaching and learning components was student questionnaires. These findings are in line with research that showed the differential validity of data sources with regard to explaining student learning (Clausen, 2001; Kunter, 2005). As a consequence, a mix of data sources and the avoidance of a single method might be fruitful for future survey studies on teaching effectiveness (Raudenbush, 2005; see also more generally, Shavelson & Towne, 2002).

Advice to the Field of Research on Teaching Effectiveness: Summary

Models of teaching and learning drive meta-analysis findings. Thus, meta-analysis researchers should be explicit about the models underlying their meta-analyses. From our point of view, the models should integrate the current state of art and allow for a comparison with past models. Meta-analyses should weight for sample size and exclude studies that do not control for student prerequisites.

Our findings have shown that the practice of reporting overall estimates of teaching effectiveness in meta-analyses is misleading. The studies going into these analyses represent heterogeneous approaches to the study of teaching effectiveness. One possible approach for future meta-analyses might be to focus on sets of studies that address the same teaching component and/or apply the same approach in studying the phenomenon. Our concern with this approach is that it would lead to a narrow view of teaching and learning. A second approach would account for the different approaches and attempt to disentangle findings. By doing so, the function of meta-analyses would change: Instead of estimating and searching for the true effect of teaching on learning, the role of meta-analysis primarily would be to capture context and outcome variation in reporting nuanced findings of teaching effectiveness.

With our meta-analysis, we sought to address researchers following both correlational and (quasi) experimental approaches: For the first group, our findings indicate that the choice and the measurement of teaching components in survey studies need to be reconsidered. For future survey studies, we suggest refining underlying theoretical models, adding teaching components proximal to executive learning activities, and applying a mix of data sources to measure teaching components. In addition, predominant statistical approaches in analyzing teaching effects on student learning should be reconsidered (Rowan et al., 2002).

With regard to quasi-experimental and experimental research, we found that the set of teaching components investigated was restricted and that researchers should be aware of the teaching components necessary for a complete cycle of learning. In addition, these researchers should expand their purview and systematically measure and investigate the effects of teaching on learning processes, motivational–affective outcomes, and cognitive outcomes.

In this meta-analysis, we have argued that effect sizes vary as to underlying research design, teaching component measurement, and type of outcome. We pointed out that (quasi) experimental research showed advantages over correlational research with regard to the specificity of the teaching component measurement. And the meta-analysis provided some evidence for an advantage of using specific measures of teaching components compared to distal and general teaching measures. However, this meta-analysis did not differentiate between the specificity of the outcome measured—that is, whether the outcome measurement was tightly linked to the topic of teaching or was a more global achievement measure. One possible explanation for the differences in effect size magnitudes between correlational and experimental research might be that outcome measures used in the latter were more closely aligned to the learning content than in the former (see Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002). Future meta-analyses should examine the sensitivity of outcome measures to what was taught and so be able to test this alternative interpretation. For that purpose, we encourage researchers to use multiple cognitive outcome measures in future meta-analyses—standardized achievement tests and content-specific tests.

APPENDIX

Publications according to teaching and learning components

Component	Publications
Time for learning	Annevelink, 2004; Arnold et al., 2002; Burkam et al., 1997; Cooper et al., 1998; Creemers & Werf, 2000; Crijnen et al., 1998; D'Agostino, 2000; De Fraine et al., 2003; de Jong et al., 2004; Driessen & Sleegers, 2000; Gamoran et al., 1997; Grift et al., 1997; Gruehn, 1995; Guthrie et al., 2001; Hill & Rowe, 1998; Hofman et al., 1999; Kunter, 2005; Kupermintz et al., 1995; Kyriakides et al., 2000; Luyten & de Jong, 1998; Martinez & Martinez, 1999; Meijnen et al., 2003; Miller & Meece, 1997; Muijs & Reynolds, 2000; Nolen, 2003; Pugh & Telhaj, 2003; Reezigt et al., 1999; Seidel et al., 2002; Taylor et al., 2003; Trautwein et al., 2002; Werf, 1997; Werf et al., 2000
Organization of learning	Baumert & Köller, 2000; Behets, 2001; Bennacer, 2000; Burkam et al., 1997; Clausen, 2001; avidson et al., 1996; Gruehn, 1995; Hamilton et al., 1995; Helmke & Weinert, 1997b; Hofman et al., 1999; Klieme & Rakoczy, 2003; Kunter, 2005; Kupermintz & Snow, 1997; Lowther et al., 2003; Luyten & de Jong, 1998; Reezigt et al., 1999; Taylor et al., 2003
Social context for learning	Anderman et al., 2001; Burkam et al., 1997; Clausen, 2001; De Fraine et al., 2003; Driessen & Sleegers, 2000; Grift et al., 1997; Gruehn, 1995; Helmke & Weinert, 1997b; Hill & Rowe, 1998; Hofman et al., 1999; Klieme & Rakoczy, 2003; Kuklinski & Weinstein, 2001; Kunter, 2005; Luyten & de Jong, 1998; Pugh & Telhaj, 2003; Staub & Stern, 2002; Stipek et al., 1998; Stolarchuk & Fisher, 2001; Van Horn & Ramey, 2003; Willms & Somers, 2001; Young, 1998
Goal setting/orientation	Alexander et al., 2002; Baumert & Köller, 2000; Chang, 2002; D'Agostino, 2000; De Fraine et al., 2003; de Jong et al., 2004; Driessen & Sleegers, 2000; Foorman et al., 1998; Grift et al., 1997; Gruehn, 1995; Helmke & Weinert, 1997b; Henderson & Landesman, 1995; Hill & Rowe, 1998; Hofman et al., 1999; Hopkins et al., 1997; Klieme & Rakoczy, 2003; Kroesbergen et al., 2004; Kunter, 2005; Kupermintz et al., 1995; Kyriakides et al., 2000; Luyten & de Jong, 1998; Marks, 2000; Muijs & Reynolds, 2000; Ramsden, 1997; Reezigt et al., 1999; Seidel et al., 2002; Seidel et al., 2005; Taylor et al., 2003; Werf, 1997; Werf et al., 2000; Yair, 2000

Execution of learning activities	
Social interaction/direct experiences	Allsopp, 1997; Applebee et al., 2003; Ashman & Gillies, 1997; Baumert & Köller, 2000; Borsch et al., 2002; Brush, 1997; Burkam et al., 1997; Clausen, 2001; Driessen & Sleegers, 2000; D. Fuchs et al., 1997; L. S. Fuchs et al., 2002; George & Kaplan, 1998; Ginsburg-Block & Fantuzzo, 1998; Gruehn, 1995; Hamilton et al., 1995; Helmke & Weinert, 1997b; Houghton et al., 1995; Jovanovic & King, 1998; Kunter, 2005; Kupermintz et al., 1995; Kupermintz & Snow, 1997; Lazarowitz et al., 1996; Mathes et al., 1998; Seidel et al., 2002; Stevens & Slavin, 1995; Stipek et al., 1998; Stolarchuk & Fisher, 2001; Sumfleth et al., 2004; Taylor et al., 2003; Van Horn & Ramey, 2003; Wenglinsky, 2002; Werf et al., 2000; Young, 1998
Basic information processing	Baumert & Köller, 2000; Burkam et al., 1997; Cantrell, 1999; Chularut & DeBacker, 2004; Clausen, 2001; D'Agostino, 2000; Einsiedler & Treinies, 1997; Gruehn, 1995; Hardy et al., 2004; Hofman et al., 1999; Klieme & Rakoczy, 2003; Kroesbergen et al., 2004; Kunter, 2005; Kupermintz et al., 1995; Kupermintz & Snow, 1997; Möller et al., 2002; Muijs & Reynolds, 2003; Newmann et al., 1996; Nolen, 2003; Pauli et al., 2003; Taylor et al., 2003; Tomoff et al., 2000; Van Horn & Ramey, 2003; Wenglinsky, 2002; Yair, 2000; Young, 1998
Domain-specific information processing	Baumann et al., 2003; Chang & Barufaldi, 1999; Driessen & Sleegers, 2000; Ginsburg-Block & Fantuzzo, 1998; Guthrie et al., 1999; Guthrie et al., 1998; Guthrie et al., 1996; Guthrie et al., 2000; Hogan, 1999; Kramarski & Mevarech, 1997; Kramarski et al., 2001; McGuinness et al., 1995; Mevarech, 1999; Mevarech & Kramarski, 2003; Perry, 1998; Taylor et al., 2003; White & Frederiksen, 1998; Wigfield et al., 2004
Evaluation of learning	Driessen & Sleegers, 2000; Gruehn, 1995; Meijnen et al., 2003; Pugh & Telhaj, 2003; Reezigt et al., 1999; Wenglinsky, 2002; Werf, 1997; Werf et al., 2000; Willms & Somers, 2001
Regulation and monitoring	Clausen, 2001; De Fraine et al., 2003; de Jong et al., 2004; Driessen & Sleegers, 2000; Grift et al., 1997; Gruehn, 1995; Hardre & Reeve, 2003; Helmke & Weinert, 1997b; Hill & Rowe, 1998; Hofman et al., 1999; Houtveen et al., 1999; Klieme & Rakoczy, 2003; Kunter, 2005; Kyriakides et al., 2000; Marks, 2000; Meijnen et al., 2003; Parcel & Dufur, 2001; Pugh & Telhaj, 2003; Reezigt et al., 1999; Seidel et al., 2003; Stipek et al., 1998; Taylor et al., 2003; Udziela, 1996; Veenstra, 1999; Werf, 1997; Werf et al., 2000; Wigfield et al., 2004; Young, 1998

Notes

We want to express our gratitude to Jaap Scheerens, who cooperated with the first author in the project Positioning the Supervision Framework 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. We also wish to thank the Dutch Educational Inspectorate for financing the database and for making the research of this article possible. Finally, we are grateful to Ingram Olkin for his advice on weighting procedures and to Manfred Prenzel for reviewing earlier versions of the article.

[1]Countries of publication are Argentina, Australia, Belgium, Bolivia, Brazil, Chile, Columbia, Costa Rica, Cuba, Cyprus, Germany, Honduras, Israel, Mexico, Netherlands, New Zealand, Norway, Paraguay, Peru, Switzerland, Dominican Republic, Taiwan, United Kingdom, United States of America, and Venezuela.

[2]Some countries have 13 years of schooling.

[3]The transformation of statistics as reported in the original articles into Fisher's *Z* as an index of effect size was part of the project Positioning the Supervision Framework 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 (Scheerens, Seidel, Witziers, Hendriks, & Doornekamp, 2005) and was carried out by Jaap Scheerens at the University of Twente, Netherlands.

[4]In addition, the analyses have been carried out for unweighted effect sizes and are available. A copy can be received by contacting the first author.

[5]Readers are reminded that effect sizes reported in the tables refer to Fisher's *Z*. To interpret effect sizes according to Cohen's *d*, use the formula presented in the Method section (as done for the interpretation of effect sizes in this paragraph).

References

References marked with an asterisk indicate studies included in the meta-analysis.

Alexander, P. A. (2000). Toward a model of academic development: Schooling and the acquisition of knowledge. *Educational Researcher, 29*(2), 28–33.

*Alexander, P. A., Fives, H., Buehl, M. M., & Mulhern, J. (2002). Teaching as persuasion. *Teaching and Teacher Education, 18*(7), 795–813.

*Allsopp, D. H. (1997). Using classwide peer tutoring to teach beginning algebra problem-solving skills in heterogeneous classrooms. *Remedial and Special Education, 18*(6), 367–379.

American Association for the Advancement of Science. (1990a). *Benchmarks for science literacy.* New York: Oxford University Press.

American Association for the Advancement of Science. (1990b). *Science for all Americans.* New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

*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, L. W. (2004). *Increasing teacher effectiveness* (2nd ed.). Paris: UNESCO, International Institute for Educational Planning.

*Annevelink, E. (2004). *Class size: Linking teaching and learning*. Unpublished doctoral thesis, University of Twente, Enschede.
*Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730.
*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.
*Ashman, A. F., & Gillies, R. M. (1997). Children's cooperative behavior and interactions in trained and untrained work groups in regular classrooms. *Journal of School Psychology, 35*(3), 261–279.
Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51,* 267–272.
*Baumann, J. F., Edwards, E. C., Boland, E. M., Olejnik, S., & Kame'enui, E. J. (2003). Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. *American Educational Research Journal, 40*(2), 447–494.
Baumert, J., Blum, W., & Neubrand, J. (2004). Drawing the lessons from PISA-2000: Long term research implications. *Zeitschrift für Erziehungswissenschaft, Beiheft 3-04,* 143–157.
*Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe [Design of instruction, insightful learning and multiple target achievement in mathematics and physics instruction in German upper level schools]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2* (pp. 271–315). Opladen, Germany: Leske + Budrich.
*Behets, C. (2001). Comparison of more and less effective teaching behaviors in secondary physical education. *Teaching and Teacher Education, 13*(2), 215–224.
*Bennacer, H. (2000). How the socioecological characteristics of the classroom affect academic achievement. *European Journal of Psychology of Education, 15*(2), 173–189.
Black, P., & Williams, N. J. (2003). "In praise of educational research": Formative assessment. *British Journal of Educational Psychology, 29*(5), 623–637.
Bloom, B. S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.
Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31,* 445–457.
Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. *Learning and Instruction, 13*(3), 327–347.
Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15.
*Borsch, F., Jurgen-Lohmann, J., & Giesen, H. (2002). Cooperative learning in elementary schools: Effects of the jigsaw method on student achievement in science. *Psychologie in Erziehung und Unterricht, 49*(3), 172–183.
Boruch, R., & Mosteller, F. (Eds.). (2002). *Evidence matters: Randomized trials in education research.* Washington, DC: Brookings Institution 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. (2000). *Teaching*. Brussels, Belgium: International Academy of Education.
Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research and teaching* (pp. 328–375). New York: Macmillan.
Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.
*Brush, T. A. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Etr&D—Educational Technology Research and Development, 45*(1), 51–64.
*Burkam, D. T., Lee, V. E., & Smerdon, B. A. (1997). Gender and science learning early in high school: Subject matter and laboratory experiences. *American Educational Research Journal, 34*(2), 297–331.
Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: R. McNally.
*Cantrell, S. C. (1999). The effects of literacy instruction on primary students' reading and writing achievement. *Reading Research and Instruction, 39*(1), 3–26.
Caroll, J. (1963). A model for school learning. *Teachers College Record, 64,* 723–733.
*Chang, C. Y. (2002). The impact of different forms of multimedia CAI on students' science achievement. *Innovations in Education and Teaching International, 39*(4), 280–288.
*Chang, C. Y., & Barufaldi, J. P. (1999). The use of a problem-solving-based instructional model in initiating change in students' achievement and alternative frameworks. *International Journal of Science Education, 21*(4), 373–388.
Chi, M. T. H., Feltovich, J. P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.
*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 [Quality of instruction. A question of the perspective? Empirical analysis of agreement, construct and criterion validity]. Münster, Germany: Waxmann.
Cohen, D. K., Raudenbush, S. W., & Loewenberg Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119–142.
Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.
Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. In R. Boruch & F. Mosteller (Eds.), *Evidence matters: Randomized trials in education research* (pp. 150–178). Washington, DC: Brookings Institution Press.
Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.
*Cooper, H., Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology, 90*(1), 70–83.

Crawford, J., & Impara, J. C. (2001). Critical issues, current trends, and possible futures in quantitative methods. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 133–173). Washington, DC: American Educational Research Association.

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

*Creemers, B. P. M., & Werf, G. v. d. (2000). Economic viewpoints in educational effectiveness: Cost-effectiveness analysis of an educational improvement project. *School Effectiveness and School Improvement, 11*(2), 197–235.

*Crijnen, A. A. M., Feehan, M., & Kellam, S. G. (1998). The course and malleability of a reading achievement in elementary school: The application of growth curve modeling in the evaluation of a mastery learning intervention. *Learning and Individual Differences, 10*(2), 137-157.

*D'Agostino, J. V. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievements. *School Effectiveness and School Improvement, 11*(2), 197–235.

*Davidson, J., Elcock, J., & Noyes, P. (1996). A preliminary study of the effect of computer-assisted practice on reading attainment. *Journal of Research in Reading, 19,* 102–110.

De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 491–549). New York: MacMillan.

De Corte, E., Verschaffel, L., Entwistle, N., & Merrienboer, v. J. (Eds.). (2003). *Powerful learning environments: Unravelling basic components and dimensions.* Amsterdam: Pergamon.

*De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal, 29*(6), 841–859.

*de Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*(1), 3–31.

Donovan, M. S., & Bransford, J. D. (Eds.). (2005). *How students learn: History, mathematics, and science in the classroom.* Washington, DC: National Academy Press.

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 392–431). New York: Macmillan.

*Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement, 11*(1), 57–79.

Dunkin, M., & Biddle, B. (1974). *The study of teaching.* New York: Holt, Rhinehart & Winston.

*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.

Floden, R. E. (2001). Research on effects of teaching: A continuing model for research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 3–16). Washington, DC: American Educational Research Association.

*Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37–55.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11,* 145–252.

*Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*(1), 174–206.

*Fuchs, L. S., Fuchs, D., Yazdian, L., & Powell, S. R. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review, 31*(4), 569–583.

Gage, N. L. (1963). Paradigms for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 94–141). Chicago: Rand McNally.

Gage, N. L., & Berliner, D. C. (1998). *Educational psychology* (6th ed.). Boston: Houghton Mifflin.

*Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325–338.

*George, R., & Kaplan, D. (1998). A structural model of parent and teacher influences on science attitudes of eighth graders: Evidence from NELS:88. *Science Education, 82*(1), 93–109.

*Ginsburg-Block, M. D., & Fantuzzo, J. W. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology, 90*(3), 560–569.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). New York: MacMillan.

*Grift, W. v. d., Houtveen, T., & Vermeulen, C. (1997). Instructional climate in Dutch secondary education. *School Effectiveness and School Improvement, 8*(4), 449–462.

*Gruehn, S. (1995). The compatibility of cognitive and noncognitive objectives of instruction. *Zeitschrift Für Padagogik, 41*(4), 531–553.

*Guthrie, J. T., Anderson, E., Alao, S., & Rinehart, J. (1999). Influences of concept-oriented reading instruction on strategy use and conceptual learning from text. *Elementary School Journal, 99*(4), 343–366.

*Guthrie, J. T., Schafer, W. D., & Huang, C. W. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research, 94*(3), 145–162.

*Guthrie, J. T., Van Meter, P., Hancock, G. R., Alao, S., Anderson, E., & McCann, A. (1998). Does concept-oriented reading instruction increase strategy use and conceptual learning from text? *Journal of Educational Psychology, 90*(2), 261–278.

*Guthrie, J. T., Van Meter, P., McCann, A. D., Wigfield, A., Bennett, L., Poundstone, C. C., et al. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly, 31*(3), 306–332.

*Guthrie, J. T., Wigfield, A., & Von Secker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92*(2), 331–341.

*Hamilton, L. S., Nussbaum, E. M., Kupermintz, H., Kerkhoven, J. I. M., & Snow, R. E. (1995). Enhancing the validity and usefulness of large-scale educational assessments: NELS:88 science achievement. *American Educational Research Journal, 32*(3), 555–581.

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*(3), 638–645.

*Hardre, P. L., & Reeve, J. (2003). A motivational model of rural students' intentions to persist in versus drop out of high school. *Journal of Educational Psychology, 95*(2), 347–356.

*Hardy, I., Jonen, A., Möller, K., & Stern, E. (2004). Die Integration von Repräsentationsformen in den Sachunterricht der Grundschule [The integration of representation forms in science and social studies in primary schools]. In J. Doll & M. Prenzel (Eds.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (pp. 267–283). Münster, Germany: Waxmann.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego: Academic Press.

Helmke, A., & Weinert, F. E. (1997a). Bedingungsfaktoren schulischer Leistungen [Conditions of school achievement]. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie: Bd. 3. Psychologie des Unterrichts und der Schule* (pp. 71–176). Göttingen, Germany: Hogrefe.

*Helmke, A., & Weinert, F. E. (1997b). Unterrichtsqualität und Leistungsentwicklung: Ergebnisse aus dem SCHOLASTIK-Projekt [Quality of instruction and development of achievement: Results from the SCHOKASTIK project]. In F. E. Weinert & A. Helmke (Eds.), *Entwicklung im Grundschlualter* (pp. 241–251).

*Henderson, R. W., & Landesman, E. M. (1995). Effects of thematically integrated mathematics instruction on students of Mexican descent. *Journal of Educational Research, 88*(5), 290–300.

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

*Hofman, R. H., Hofman, W. H. A., & Guldemond, H. (1999). Social and cognitive outcomes: A comparison of contexts of learning. *School Effectiveness and School Improvement, 10*(3), 352–366.

*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, K. B., McGillicuddy-DeLisi, A. V., & DeLisi, R. (1997). Student gender and teaching methods as sources of variability in children's computational arithmetic performance. *Journal of Genetic Psychology, 158*(3), 333–345.

*Houghton, S., Litwin, M., & Carroll, A. (1995). Peer mediated intervention in reading instruction using pause, prompt and praise. *Educational Studies, 21*(3), 361–377.

*Houtveen, A. A. M., Booij, N., de Jong, R., & van de Grift, W. (1999). Adaptive instruction and pupil achievement. *School Effectiveness and School Improvement, 10*(2), 172–192.

*Jovanovic, J., & King, S. S. (1998). Boys and girls in the performance-based science classroom: Who's doing the performing? *American Educational Research Journal, 35*(3), 477–496.

Kardash, C. M., & Amlund, J. T. (1991). Self-reported learning strategies and learning from expository text. *Contemporary Educational Psychology, 16,* 117–138.

*Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht [Instructional quality from the students' perspective: Culture-specific profiles, regional differences, and relationships with the effects of instruction]. 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, Germany: Leske + Budrich.

*Kramarski, B., & Mevarech, Z. R. (1997). Cognitive-metacognitive training within a problem-solving based Logo environment. *British Journal of Educational Psychology, 67,* 425–445.

*Kramarski, B., Mevarech, Z. R., & Lieberman, A. (2001). Effects of multilevel versus unilevel metacognitive training on mathematical reasoning. *Journal of Educational Research, 94*(5), 292–300.

*Kroesbergen, E. H., Van Luit, J. E. H., & Maas, C. J. M. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *Elementary School Journal, 104*(3), 233–251.

*Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development, 72*(5), 1554–1578.

*Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster, Germany: Waxmann.

*Kupermintz, H., Ennis, M. M., Hamilton, L. S., Talbert, J. E., & Snow, R. E. (1995). Enhancing the validity and usefulness of large-scale educational assessments. NELS:88 mathematics achievement. *American Educational Research Journal, 32*(3), 525–554.

*Kupermintz, H., & Snow, R. E. (1997). Enhancing the validity and usefulness of large-scale educational assessments. NELS:88 mathematics achievement to 12th grade. *American Educational Research Journal, 34*(1), 124–150.

*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.

*Lazarowitz, R., Baird, J. H., & Bowlden, V. (1996). Teaching biology in a group mastery learning mode: High school students' academic achievement and affective outcomes. *International Journal of Science Education, 18*(4), 447–462.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist, 48*(12), 1181–1209.

*Lowther, D. L., Ross, S. M., & Morrison, G. M. (2003). When each one has one: The influences on teaching strategies and student achievement of using laptops in the classroom. *Etr&D—Educational Technology Research and Development, 51*(3), 23–44.

*Luyten, H., & de Jong, R. (1998). Parallel classes: Differences and similarities. Teacher effects and school effects in secondary schools. *School Effectiveness and School Improvement, 9*(4), 437–473.

*Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153–184.

*Martinez, J. G. R., & Martinez, N. C. (1999). Teacher effectiveness and learning for mastery. *Journal of Educational Research, 92*(5), 279–285.

*Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33*(1), 62–94.

*McGuinness, D., McGuinness, C., & Donohue, J. (1995). Phonological training and the alphabet principle: Evidence for reciprocal causality. *Reading Research Quarterly, 30*(4), 830–852.

*Meijnen, G. W., Lagerweij, N. W., & Jong, P. F. (2003). Instruction characteristics and cognitive achievement. *School Effectiveness and School Improvement, 14*(2), 159–187.

*Mevarech, Z. R. (1999). Effects of metacognitive training embedded in cooperative settings on mathematical problem solving. *Journal of Educational Research, 92*(4), 195–205.

*Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2), 365–394.

*Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology, 73,* 449–471.

*Miller, S. D., & Meece, J. L. (1997). Enhancing elementary students' motivation to read and write: A classroom intervention study. *Journal of Educational Research, 90*(5), 286–299.

*Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. *Zeitschrift für Pädagogik, 45. Beiheft,* 176–191.

*Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: Some preliminary findings from the evaluation of the mathematics enhancement programme (primary). *School Effectiveness and School Improvement, 11*(3), 273–303.

*Muijs, D., & Reynolds, D. (2003). The effectiveness of the use of learning support assistants in improving the mathematics achievement of low achieving pupils in primary school. *Educational Research, 45*(3), 219–230.

National Research Council. (1990). *Reshaping school mathematics: A philosophy and framework for curriculum.* Washington, DC: National Academy Press.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council. (2001). *Inquiry and the National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council. (2004). *Implementing randomized field trials in education: Report of a workshop.* Washington, DC: National Academy Press.

National Research Council. (2005). *Advancing scientific research in education.* Washington, DC: National Academy Press.

*Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*(4), 280–312.

*Nolen, S. B. (2003). Learning environment, motivation, and achievement in high school science. *Journal of Research in Science Teaching, 40*(4), 347–368.

OECD. (2004). PISA 2006 Scientific Literacy Framework. Paris: Author.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension monitoring activities. *Cognition and Instruction, 1,* 117–175.

*Parcel, T. L., & Dufur, M. J. (2001). Capital at home and at school: Effects on student achievement. *Social Forces, 79*(3), 881–911.

*Pauli, C., Reusser, K., Waldis, M., & Grob, U. (2003). "Enrichment of learning settings and teaching methods" in mathematics lessons in the German-speaking part of Switzerland. *Unterrichtswissenschaft, 31*(4), 291–320.

*Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology, 90*(4), 715–729.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33–40.

*Pugh, G., & Telhaj, S. (2003, September). *Attainment effects of school enmeshment with external communities: Community policy, church/religious influence, and TIMSS–R mathematics scores in Flemish secondary schools*. Paper presented at the European Conference on Educational Research, Hamburg, Germany.

*Ramsden, J. M. (1997). How does a context-based approach influence understanding of key chemical ideas at 16+? *International Journal of Science Education, 19*(6), 697–710.

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioural Statistics, 29*(1), 121–129.

Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher, 34*(5), 25–31.

*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.

Reusser, K., & Stebler, R. (1997). Every word problem has a solution: The social rationality of mathematical modeling in schools. *Learning and Instruction, 7*(4), 309–327.

Rosenshine, B. (1979). Content, time, and direct instruction. In P. Peterson & H. J. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications.* Berkeley, CA: McCutchan.

Rosenthal, R. (1994). 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., Stringfield, S., Sanders, W. L., & Wright, S. P. (2003). Inside systemic elementary school reform: Teacher effects and teacher mobility. *School Effectiveness and School Improvement, 14*(1), 73–110.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*(8), 1525–1567.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5), 369–393.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54–67.

Scheerens, J. (2000). *Improving school effectiveness.* Paris: UNESCO.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness.* Oxford, UK: Pergamon.

Scheerens, J., Seidel, T., Witziers, B., Hendriks, M., & Doornekamp, G. (2005). *Positioning the supervision framework 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, Netherlands/Kiel, Germany: University of Twente/Institute for Science Education (IPN).

Schoenfeld, A. H. (1985). *Mathematical problem solving.* London: Academic Press.

*Seidel, T., Prenzel, M., Duit, R., Euler, M., Geiser, H., Hoffmann, L., et al. (2002). "Can everybody look to the front of the classroom please?"—Patterns of instruction in elementary physics classrooms and its implications for students' learning. *Unterrichtswissenschaft, 30*(1), 52–77.

*Seidel, T., Rimmele, R., & Prenzel, M. (2003). Opportunities for learning motivation in classroom discourse—Combination of video analysis and student questionnaires. *Unterrichtswissenschaft, 31*(2), 142–165.

*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.

Shadish, W. R., Campbell, D. T., & Cook, T. D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25–28.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education.* Washington, DC: National Academy Press.

Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teaching. In M. C. Wittrock (Ed.), *Handbook of research and teaching* (pp. 50–91). New York: Macmillan.

Shuell, T. J. (1993). Toward an integrated theory of teaching and learning. *Educational Psychologist, 28,* 291–311.

Shuell, T. J. (1996). Teaching and learning in a classroom context. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York: Macmillan.

Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 3–36). New York: Macmillan.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

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

Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243–310). New York: MacMillan.

Snow, R. E., Frederico, P.-A., & Montague, W. E. (Eds.). (1980). *Aptitude, learning, and instruction.* Hillsdale, NJ: Lawrence Erlbaum.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology, 43,* 583–626.

*Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94*(2), 344–355.

*Stevens, R. J., & Slavin, R. E. (1995). Effects of a cooperative learning approach in reading and writing on academically handicapped and nonhandicapped students. *Elementary School Journal, 95*(3), 241–262.

*Stipek, D., Salmon, J. M., Givvin, K. B., Kazemi, E., Saxe, G., & MacGyvers, V. L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education, 29*(4), 465–488.

*Stolarchuk, E., & Fisher, D. (2001). An investigation of teacher-student interpersonal behavior in science classrooms using laptop computers. *Journal of Educational Computing Research, 24*(1), 41–55.

*Sumfleth, E., Rumann, S., & Nicolai, N. (2004). Schulische und häusliche Kooperation im Chemieanfangsunterricht [Cooperation at the school and at home in elementary chemistry instruction]. In J. Doll & M. Prenzel (Eds.), *Bildungsqualität von Schule:*

Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung (pp. 284–302). Münster, Germany: Waxmann.
*Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal, 104*(1), 3–28.
*Tomoff, J., Thompson, M., & Behrens, J. (2000). *Measuring NCTM-recommended practices and student achievement with TIMSS.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
*Trautwein, U., Köller, O., Schmitz, B., & Baumert, J. (2002). Do homework assignments enhance achievement? A multilevel analysis in 7th-grade mathematics. *Contemporary Educational Psychology, 27*(1), 26–50.
*Udziela, T. (1996). *Effect of formal study skills training on sixth grade reading achievement* (Reports–Research/Technical, Chicago suburbs). Illinois Goal Assessment Program.
*Van Horn, M. L., & Ramey, S. L. (2003). The effects of developmentally appropriate practices on academic outcomes among former head start students and classmates, Grades 1-3. *American Educational Research Journal, 40*(4), 961–990.
*Veenstra, D. R. (1999). *Leerlingen, klassen, scholen*. Unpublished doctoral thesis. Thela Thesis, Amsterdam.
Walberg, H. J. (1986). Syntheses of research on teaching. In M. C. Wittrock (Ed.), *Handbook of research and teaching* (pp. 214–229). New York: Macmillan.
Walberg, H. J., & Paik, S. J. (2000). *Effective educational practices.* Brussels, Belgium: International Academy of Education.
Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63,* 249–294.
Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122.
*Wenglinsky, H. (2002). How school matter: The link between teacher classroom practices and student academic performance [Online]. *Education Policy Analysis Archives, 10*(12).
*Werf, G. v. d. (1997). Differences in school and instruction characteristics between high-, average-, and low-effective schools. *School Effectiveness and School Improvement, 8*(4), 430–448.
*Werf, G. v. d., Creemers, B. P. M., Jong, R., & Klaver, E. (2000). Evaluation of school improvement through an educational effectiveness model: The case of Indonesia's PEQIP project. *Comparative Education Review, 44*(3), 329–355.
*White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.
*Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading: Domain specificity and instructional influences. *Journal of Educational Research, 97*(6), 299–309.
*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.
*Yair, G. (2000). Reforming motivation: How the structure of instruction affects students' learning experiences. *British Educational Research Journal, 26*(2), 191–210.
*Young, D. Y. (1998). Rural and urban differences in student achievement in science and mathematics: A multilevel analysis. *School Effectiveness and School Improvement, 9*(4), 386–418

Authors

TINA SEIDEL was an assistant professor of education at the Leibniz-Institute for Science Education (IPN) at the University of Kiel, Olshausenstr. 62, 24098 Kiel, Germany. Since April 2007 she has been a professor of educational psychology at Friedrich Schiller University Jena, Am Planetarium 4, 07743 Jena, Germany; e-mail: *tina.seidel@uni-jena.de*. Her research interests are in processes of classroom teaching and learning, teacher education, and higher education.

RICHARD J. SHAVELSON is the Margaret Jack Professor of Education, Professor of Psychology (courtesy), and former I. James Quillen Dean of the School of Education at Stanford University and Senior Fellow in the Woods Institute for the Environment at Stanford, 485 Lasuen Mall, Stanford, CA 94305-3096, USA; e-mail: *richs@stanford.edu.* His current work includes the assessment of science achievement and the study of inquiry-based science teaching and its impact on students' knowledge structures and performance. Other current work includes (a) studies of the causal impact of computer cognitive training on working memory, fluid intelligence and science achievement; (b) assessment of undergraduates' learning with the Collegiate Learning Assessment; and (c) accountability in higher education.