9

THE EFFECTS OF CLASSROOM MATHEMATICS TEACHING ON STUDENTS' LEARNING

James Hiebert

UNIVERSITY OF DELAWARE

Douglas A. Grouws

UNIVERSITY OF MISSOURI

TEACHING MATTERS . . . BUT *HOW* IS NOT EASY TO DOCUMENT

We begin with the following claim: The nature of classroom mathematics teaching significantly affects the nature and level of students' learning. Such a claim seems obvious. Everyone can remember teachers who were especially effective. Surely, teaching makes a difference. Systematically collected data support most people's personal experience by showing that different teachers produce different levels of student achievement gains (Nye, Konstantopoulos, & Hedges, 2004; Sanders & Rivers, 1996; Sanders, Saxton, & Horn, 1997). In fact, the cumulative effect over several years of effective teachers is substantial. Having good teachers really does make a difference.

But what makes mathematics teachers effective? This question does not have an obvious or easy answer. The answer is not found by searching the reports cited above on teacher effectiveness. Many reasons could explain why some of the teachers in these studies were more effective than others. Classrooms are filled with complex dynamics, and many factors could be responsible for increased student learning. This is not one of those pedantic educational issues that researchers get tangled up in by splitting hairs and that, in the end, simply demonstrate what everyone already knows. This is a very central and difficult question to answer. It is, in a real sense, the core question of education. What is it, exactly, about teaching that influences students' learning? Although laypersons in the street often have quick answers based on recollections of their favorite teachers, documenting particular features of teaching that are consistently effective for students' learning has proven to be one of the great research challenges in education.

Our aim in this chapter is to tackle directly the issue of teaching effectiveness—why it has been so hard to document, what is known about it, and how the mathematics education community can learn more. We first examine why it has been so difficult to establish robust links between teaching and learning, then we present a few claims that organize and structure the literature on teaching effects in what we hope are helpful ways, and finally we outline a set of goals and strategies to guide future work in this central and urgent research domain. Although we believe the claims we make about links between mathematics teaching and learning are well supported, we do not present an exhaustive review of the research literature to substantiate them. Rather, we develop arguments to support our claims, select a few studies that illustrate them, and refer the reader to other, more extensive reviews of relevant research.

Before considering some of the difficulties facing researchers who try to document the effects of teaching, we need to define what we mean by *teaching*. Simple definitions, presented in some classic treatments of teaching, are a good starting point. Thorndike (1906) defined *teaching* as the methods used to help students achieve the learning goals valued by society. Gage (1978) defined *teaching* as "any activity on the part of one person intended to facilitate learning by another" (p. 14). These definitions are incomplete because, although they contain much of what we mean by *teaching*, they treat teaching as a one-way relationship—teachers acting on students. In reality, teaching is influenced by students and has a bidirectional quality. Cohen, Raudenbush, and Ball (2003) captured this quality by saying that "instruction consists of interactions among teachers and students around content" (p. 122). If we use Cohen et al.'s definition but add back the goal-directedness explicit in Thorndike's definition and restrict the environments to school classrooms, we have a working definition that sets the boundaries for this chapter: Teaching consists of classroom interactions among teachers and students around content directed toward facilitating students' achievement of learning goals.

Why has it been so difficult to identify features of teaching, as defined here, that show consistent and strong effects on students' learning? Examining this question sets the context for both the claims we make about links between mathematics teaching and learning and the strategies we propose to advance this central research domain in mathematics education. Because the issues surrounding the difficulty of linking teaching with learning are not unique to mathematics, and because they have been addressed in the broader literature, we examine the question both within this broader context and within contexts specific to mathematics.

A Claim That Appears Obvious is Strikingly Difficult to Specify

Despite more than 100 years of work on specifying effective teaching, the field still is searching for reliable, demonstrable recommendations of teaching methods. This is not to say that recommendations have not been proposed. The education literature is filled with suggestions, recommendations, and sure-to-succeed descriptions of teaching. Pestalozzi, one of the great educators of the past, was said to have claimed discovery of "the ultimate method of effective instruction" in the early 1800s (Reusser, 2001, p. 1). A perusal of more recent analyses indicates that Pestalozzi's optimism is alive but not widespread.

Reviews of the teaching-learning literature indicate considerable differences of opinion, not just about which teaching methods are effective, but even about how much is known about whether any methods are particularly effective. Some researchers portray a field that is making good progress in documenting links between teaching and learning, but other researchers see less progress. The third edition of the *Handbook of Research on Teaching* includes two chapters that review, with considerable confidence, the connections between particular teaching behaviors and students' learning (Brophy & Good, 1986; Rosenshine & Stevens, 1986). More recently, Brophy (1999) reinforced this optimism by identifying 12 principles of effective teaching based on accumulated research.

In contrast to these optimistic claims, Romberg and Carpenter (1986) argued that research on teaching and research on learning had constituted two separate fields in mathematics education with few attempts to connect learning and teaching. Consequently, there was little to say about specific links between the two. Romberg and Carpenter's less optimistic view is found in the same third edition of the *Handbook* that contains the positive treatments by Brophy and Good (1986) and Rosenshine and Stevens (1986).

Observations about the absence of evidence-based connections between teaching and learning can be traced back to at least the middle of the last century. A report released by the American Educational Research Association Committee on the Criteria of Teacher Effectiveness concluded

> The simple fact of the matter is that, after forty years of research on teacher effectiveness during which a vast number of studies have been carried out, one can point to few outcomes that a superintendent of schools can safely employ in hiring a teacher or granting him [*sic*] tenure, that an agency can employ in certifying teachers, or that a teacher education faculty can employ in planning or improving teacher education programs. (Barr et al., 1952, cited in Duffy, 1981, p. 113)

Two decades later, several reviews suggested that efforts to link particular teaching features with students' learning still had produced few results (Berliner, 1976; Heath & Nielson, 1974).

Even with continuing attention to the effects of classroom teaching on learning since these early reviews, researchers still are not uniformly persuaded by the evidence (Shuell, 1996). Countering Brophy's (1999) optimistic view, the two chapters in the most recent fourth edition of the *Handbook of Research on Teaching* that deal explicitly with the effects of teaching on learning are decidedly cautious about claims that link teaching with learning (Floden, 2001; Oser & Baeriswyl, 2001).

The uncertainty in the research community about what is known regarding the effects of teaching on learning can be explained by some of the challenges facing researchers engaged in this work and the ways in which the challenges have (and have not) been met. Before examining these challenges, we must point out that this uncertainty in the field has its costs. Among them is the perception by key U.S. policymakers that research does not inform classroom practice:

> Education has not relied very much on evidence, whether in regard to how to train teachers, what sort of curriculum to use or what sort of teaching methods to use. The decisions have been based on professional wisdom or the spirit of the moment rather than on research (Grover Whitehurst, U.S. Assistant Secretary of Education, as quoted by J. Traub in the *New York Times*, Nov. 10, 2002).

Pressures are increasing to provide evidence-based descriptions of effective mathematics teaching. Consequently, the focal issue of this chapter is not only fundamental, from a research perspective, but urgent, from a policy perspective.

The first step in making further progress in establishing connections between teaching and learning is understanding why this is so hard to do. In the following sections, we outline a number of difficulties that face researchers who try to connect teaching with learning. These are not excuses; they are severe obstacles that must be resolved. Indeed, after studying the following challenges, we expect that readers might change their question from "Why is it so hard to connect teaching with learning?" to "Is it even possible to establish reliable connections between particular features or methods of teaching with students' learning?" Anticipating this discouragement, we continue the chapter by developing several claims about features of teaching that affect learning, claims that we believe have accumulated considerable empirical support in spite of the challenges.

Useful Theories of Teaching Are in Short Supply

Making progress in any scientific field is difficult without explicit theories. In fact, some researchers would argue that fields are scientific to the extent that empirical research is guided by theory (National Research Council, 2002). Theories are useful because they direct researchers' attention to particular relationships, provide meaning for the phenomena being studied, rate the relative importance of the research questions being asked, and place findings from individual studies within a larger context. Theories suggest where to look when formulating the next research questions and provide an organizational scheme, or a story line, within which to accumulate and fit together individual sets of results. Karmiloff-Smith and Inhelder (1974) titled a provocative article, "If You Want to Get Ahead, Get a Theory." They were describing children's efforts to develop their understanding of physical phenomena, but the analogy holds equally well for researchers developing an understanding of connections between teaching and learning.

Robust, useful theories of classroom teaching do not yet exist. Theories that consider connections between classroom teaching and students' learning are even less developed (Floden, 2001; Oser & Baeriswyl, 2001). Within mathematics, theories of learning have been more clearly articulated than theories of teaching. Although theories of learning provide some guidance for research on teaching, they do not translate directly into theories of teaching and cannot be used, by themselves, to stimulate or coordinate research agendas on the effects of teaching on learning. Components or features of teaching that are likely to play important roles in useful working theories have been identified, and beginnings of theories of mathematics teaching have been proposed (e.g., Brousseau, 1997; Freudenthal, 1973; Gravemeijer, 1994; Leinhardt & Greeno, 1986; National Research Council, 2001; Schoenfeld, 1998; Simon, Tzur, Heinz, & Kinzel, 2004). These efforts signal important progress. But theories that specify the ways in which the key components of teaching fit together to form an interactive, dynamic system for achieving particular learning goals have not been sufficiently developed to guide research efforts that can build over time.

Without guiding theories, researchers are feeling their way through murky waters. Studies that have focused on teaching-learning relationships often are chosen because they align with the personal interests of researchers. When researchers pursue different research questions and use different measures of teaching and of learning, their findings often remain isolated from one another. The data do not easily accumulate to reveal patterns and general principles. To surmount these obstacles and move ahead, researchers need to construct useful theories. Those who attempt to develop theories face at least three major conceptual challenges.

Challenge 1: Different Teaching Methods Might Be Effective for Different Learning Goals

There is no reason to believe, based on empirical findings or theoretical arguments, that a single method of teaching is the most effective for achieving all types of learning goals. Perhaps some methods of teaching are more effective for, say, memorizing number facts whereas other methods of teaching are more effective for deepening conceptual understanding and still other methods are more effective for acquiring smooth execution of complex procedures. In addition, some methods of teaching might be especially effective for showing gains in the short-term whereas others might be better for retaining and increasing gains in the long-term. Because a range of goals might be included in a single lesson, and almost certainly in a multi-lesson unit, the best or most effective teaching method might be a mix of methods, with timely and nimble shifting among them.

One consequence of recognizing that different teaching methods might be effective for different learning goals is the fact that empirical studies that compare one teaching method to another using blunt learning measures are difficult to interpret. For what kinds of learning are the teaching methods claimed to be effective?

Challenge 2: Teaching is a System of Interacting Features

A teaching method consists of multiple features that interact with one another in almost countless ways (Cohen et al., 2003; Design-Based Research Collective, 2003; Stigler & Hiebert, 1999). Theories of teaching must account for this fact by treating teaching as a system of interacting features rather than a collection of independent and interchangeable features. This means that the effects of features of teaching on students' learning cannot be measured independently of the system in which they operate. The effects of a particular feature on students' learning likely depend on its interactions with other features in the system.

Two simple examples illustrate the system nature of teaching. One of the striking results from the 1995 Video Study of the Third International Mathematics and Science Study (TIMSS) was the difference in teaching eighth-grade mathematics found between Japan and the United States. In general, lessons in Japan challenged students to solve difficult mathematics problems whereas lessons in the United States emphasized the practice of lower level skills (Stigler & Hiebert, 1999). Casual observation of the lessons suggests that Japanese teachers asked more higher level questions than U.S. teachers, but a closer look showed that the major difference was in the *kinds* of higher order questions that were asked and *when* the questions were asked (Kawanaka & Stigler, 1999). Japanese teachers asked most of their higher order questions about student-presented solution methods for solving problems whereas U.S. teachers asked most of their higher order questions about methods they, themselves, were demonstrating. Japanese teaching was distinguished not just by the existence of high-level questions but in how the questions fit within and reinforced the system of teaching being employed.

A second example of the system nature of teaching comes from research on a feature of teaching often associated with current recommendations—asking students to work in small groups. Clearly whether small groups function productively to help students achieve learning goals depends on many surrounding factors, including the knowledge and skill students acquire for working collaboratively and the kinds of tasks they are assigned (Good, Mulryan, & McCaslin, 1992; Webb, 1991; Webb, Troper, & Fall, 1995). While the first author of this chapter was reviewing videotaped lessons as part of the TIMSS, he observed one eighth-grade teacher asking students to break into their small groups and work together to answer the question "What is the name for a 12-sided object?" Students quizzed each other quickly on whether anyone knew the name and then visited for the remaining time about non-mathematical topics. This task did not lend itself to collaborative investigation. The effect of small-group work is shaped by the system of which it is a part.

A consequence of the system nature of teaching is that the effects on students' learning of individual features of teaching, like higher-order questions or the use of small groups, are difficult to isolate. Systems of teaching, not single features of teaching, affect students' learning. This is not to say that all features within a system are equally responsible for their effects on student learning. Particular features, or combinations of features, might account for more of the system's effects than other features. But the effect of any particular feature is influenced by related and interacting features. Theories of teaching must account for this fundamental complexity.

Challenge 3: The Influence of Mediating Variables

The effects of teaching on learning are complicated, not just by the features of teaching interacting with each other, but by factors that mediate the effects of the system. In a classic analysis, Wittrock (1986) argued that teaching does not affect learning directly. The effects of teaching are mediated by students' thinking—their attentiveness during instruction, their interpretations of the teacher's presentations and of the tasks they are assigned, their entry knowledge and skills, and so on. Wittrock was not the first to comment on the important role that students play in determining the effects of teaching on learning (Berliner, 1976; Doyle, 1978; Winne & Marx, 1980). Berliner (1976), for example, cited "a student's active time-on-task" (p. 10) as a complicating factor in connecting teaching and learning. More recently, Weinert, Schrader, and Helmke (1989) proposed an even stronger claim by saying that "the activity and cognitive orientation of the learners are as important as the activities and the behaviors of the teachers" (p. 899) when specifying the relationships between teaching and learning.

The set of variables related to students' thinking is not the only factor that mediates the effects of teaching on learning. A large number of contextual conditions undoubtedly influence the nature and level of students' learning (Berliner, 1976; Dunkin & Biddle, 1974). These include conditions both inside and outside of the classroom—number of students in class, length of daily lessons, home support for completing homework, and so on.

Some of the variables that mediate the effects of teaching on learning exert a stronger influence on the relationship than simply redirecting or moderating the impact of teaching on learning. These variables can reverse the causal flow so that they change the nature of teaching. For example, students can push back and alter the teacher's intended approach in both subtle and not-so-subtle ways (Cohen et al., 2003; Cooney, 1985). This is why we adjusted Thorndike's and Gage's definitions of *teaching*, as described earlier, to include two-way interactions rather than just one-way interactions from the teacher to the students.

Developing theories of teaching and, in particular, theories that connect teaching with learning is clearly a challenging task. The lack of robust widely-accepted theories is not surprising. Although the development of such theories would enable a more quickly accumulating research base, researchers would not be wise to wait for such theories to pursue empirical studies. Theoretical and empirical work can boost each other. Unfortunately, as we argue in the next section, conducting empirical studies on teaching-learning connections is fraught with a host of additional challenges. As the reader will notice, however, not all of these challenges are unique to empirical research; some are related to challenges we have already identified that confront theory development.

Documenting the Effects of Teaching on Learning Is Methodologically Difficult

Collecting trustworthy data on the effects of teaching on learning is hampered by a number of methodological difficulties. We place three kinds of difficulties at the top of the list for posing critical threats to research efforts and accounting for the challenge of linking teaching and learning. In Gage's (1978) terms, the first considers what variables to study and the second asks how to measure them. The third kind of difficulty—individual differences among students—has reappeared in various forms throughout the history of research on teaching and learning and remains a significant issue.

Accounting for Relevant Factors

Many, many factors, inside and outside of school, influence what and how well students learn. In addition to quality of teaching, influencing factors include things like the level of home support (health care, education values and resources, and so forth), peer pressures and social relationships, students' motivation, curricular materials, and assessments. To determine the effects of teaching in the midst of all the competing forces, researchers need to account for these forces in some way. Ideally, researchers would conduct experiments in which all of these forces are equalized across comparison groups, or at least are randomly distributed across the groups. Given the nature of school constraints, randomly assigning students to treatments for sufficiently long periods of time is often impossible. The consequence is that the results of studies can have many alternative interpretations. Numerous factors, other than teaching methods or in combinations with them, can explain the learning outcomes.

If researchers are able to conduct studies with random assignment of students to teaching methods, a problem remains of implementing the same method in a consistent way across classrooms. As we have noted, teaching interacts with students (as well as other contextual variables, such as the teacher's experience and expertise) and is likely to change somewhat from classroom to classroom. It can be difficult to know whether the critical features of teaching are being implemented in the same way across all sites. Limiting the research design to a single classroom or a single teacher does not solve the problem because it leaves to later replications the question of whether the findings will apply to other classrooms and teachers.

A final thorny issue arises when testing the relative benefits of particular features of teaching. As noted earlier, teaching is a system of interacting elements. Isolating particular features of the system is difficult and can even distort the effects of teaching being compared. As an example, consider the question of whether concrete manipulative materials aid students' learning.[1] First, one would need to decide how students' learning would be measured. This, in itself, is a challenging issue because different kinds of learning might be possible with manipulative materials than without such materials. Suppose this issue was resolved and treatments were designed with and without materials. Should the treatments contain the same teaching features except for the use of materials? Or should two different treatments be designed—one to provide the most effective possible approach with materials and another without materials? The first choice might lead to rather contrived treatments for one or both conditions because, under normal conditions, it is unlikely that teaching methods would be identical except for the presence or absence of materials. The second choice would lead to results that would be difficult to interpret—between-group differences might be due to the use of materials or, instead, to other differences between the methods that arose as a consequence of changing other features in the teaching system.

Accounting for all the relevant factors when designing studies or interpreting results on relationships between teaching and learning is a serious methodological challenge. As we will argue later, this does not mean that planned, comparative research is impossible, but it does suggest that the simple isolation of variables within teaching systems is unrealistic.

Creating Appropriate Measures

Developing instruments to measure students' learning and instruments to describe classroom teaching that are reliable *and* valid is a continuing methodological challenge (Ball & Rowan, 2004; Berliner, 1976), especially when using the instruments with large samples (Hamilton et al., 2003; Rowan, Correnti, & Miller, 2002). Describing teaching is, in many ways, more challenging than measuring students' learning due, in part, to its bewildering complexity and, in part, to the relatively less attention it has received.

Attempts to describe classroom teaching must consider what to measure and how to measure it. The question of what to measure is tied closely to our earlier discussions of developing theories that point researchers to important features of teaching and discussions of accounting for the relevant factors in the teaching system. We focus here on the issue of *how* to describe teaching, once the "what" is determined.

Two methods for describing teaching have been used most heavily: classroom observations (Dunkin & Biddle, 1974) and survey questionnaires (Rowan et al., 2002). Both methods suffer from limitations that must be accounted for when designing studies of teaching. Classroom observations can be influenced by biases of the observer. Establishing and maintaining inter-observer reliability on subtle but potentially important features of teaching is difficult. Observation protocols can be worked out in advance to reduce observer bias, but this requires deciding in advance what to observe. Deciding what to observe, in turn, presumes theories (at least hypotheses) of relationships between teaching and learning. Teacher questionnaires have even more serious problems, especially when the questionnaires are administered on a single occasion and ask teachers to make judgments about teaching approaches and emphases (Hiebert & Stigler, 2000; Malara & Zan, 2002; Mayer, 1999; Wubbels, Brekelmans, & Hooymayers, 1992). Teachers have a tendency to over-report features of teaching that they believe are expected. In other words, self-report surveys might indicate how teachers think they should be teaching, but the validity of these reports is suspect for describing teaching practices.

Alternative instruments to describe teaching are being developed. Video provides numerous advantages over observations and questionnaires (Stigler, Gallimore, & Hiebert, 2000) both in terms of reliability and validity. In addition, it allows teaching to be analyzed by different researchers using different approaches over long periods of time, and it affords serendipitous discoveries. When researchers can afford the expense of collecting and analyzing videotape samples of teaching, it offers many methodological benefits. Less expensive alternatives that can be used routinely on a large scale include redesigned and carefully targeted questionnaires (Ross, McDougall, Hogaboam-Gray, & LeSage, 2003; Rowan et al., 2002), combinations of measures such as questionnaires and case studies (Stecher & Borko, 2002), teacher logs that include frequent diary-like notes by teachers (Porter, Floden, Freeman, Schmidt, & Schwille, 1988; Rowan, Harrison, & Hayes, 2004), and classroom artifacts that include information regarding teachers' use of instructional materials and teaching methods, classroom activities, and students' work (Borko, Stecher,

[1] See Hiebert (2003) for a similar analysis with respect to the effects of using calculators on students' computational skill.

Alonzo, Moncure, & McClam, 2003). Many of these instruments are currently in the development and testing phase, so it is too early to evaluate their ultimate usefulness.

One methodological issue that cuts across questions of what to measure and how to measure it is the unit of analysis. What constitutes a meaningful unit of teaching? Much of the research that focused on connections between specific teaching features and student outcomes, often referred to as process-product research (Dunkin & Biddle, 1974), used relatively narrow or local units, such as the type of questions a teacher asked or the amount of time spent on each instructional task. Because we believe that teaching is best viewed as a system of interacting features, we propose using broader units that preserve the potentially important interactions. The typical daily lesson is one such unit (Stigler & Hiebert, 1999). The daily lesson has the advantage of being large enough to include key interactions among teaching features and small enough to be thoroughly analyzable. A disadvantage of using the daily lesson as the unit is that students' learning does not occur in lesson chunks. Sequences of lessons defined by full treatment of particular concepts or topics constitute other possible units of teaching, but these can become too large to digest and analyze in detail. Clearly, the tradeoffs must be weighed carefully.

Teaching Often Gets Interpreted as Teachers

To conclude the discussion of why the obvious sounding claim of "teaching affects learning" is so difficult to specify, we focus on a conceptual issue. Questions of how teaching affects learning often get muddled with questions of how teachers affect learning. Some might view this distinction as less substantive and more pedantic than the theoretical and methodological challenges posed in the earlier sections, but we believe the confusion between teachers and teaching is important because it can interfere with theoretical and methodological efforts to establish clearer connections between teaching and learning. If the differences between teachers and teaching are not clarified, research findings are misinterpreted and misunderstood.

To reiterate, we focus on teach*ing* in this chapter—classroom interactions among teachers and students around content directed toward facilitating students' achievement of learning goals. This is different from teach*ers*—their beliefs, subject matter knowledge, years of experience, personality traits, and so on. Characteristics of teachers surely can influence their teaching, but these characteristics *do not determine* their teaching. Teachers with different characteristics can teach in essentially the same way and vice versa. No doubt this a major reason for the absence of clear links between teachers' personal and background characteristics and student achievement (Dunkin & Biddle, 1974).

We chose to focus on teaching in this chapter, with a targeted and constrained definition, not only to make our task manageable but also because the teaching we define is largely under the control of the teacher. Some characteristics of teachers, such as personality traits, are quite unalterable, but the methods used to teach can be changed. This is not to say that changing teaching is easy, but it is something that teachers' control. Teaching is something that can be altered under supportive conditions and that can benefit from research-based information.

The distinction between teachers and teaching has been recognized by researchers for some time, but the distinction often disappears. In their classic 1974 book, Dunkin and Biddle recognized the distinction and even gave teacher characteristics a special name—*presage* variables. But the distinction has not always been made, perhaps because of a mistaken assumption that teacher characteristics determine, or nearly determine, the nature of teaching. The confusion is especially apparent, and nearly pervasive, in the public press and in many policy documents. Most news reports and policy statements, for example, that critique education and call for better teaching point to characteristics of teachers, not methods of instruction. The annual report on the condition of U.S. education sponsored by the Pew Charitable Trust has a section on "Teaching Quality" but reports no information on teaching methods, only characteristics of teachers such as certification programs and years of experience (Olson, 1997). In the first two sentences of the 2004 annual report of the U.S. Secretary of Education, *teacher* and *teaching* are equated and then used interchangeably with improved teaching apparently assumed to result from changing teacher characteristics (Paige, 2004). The policy document *Before It's Too Late* produced by the highly visible commission chaired by U.S. Senator John Glenn includes four key points in the foreword about improving mathematics and science teaching in the 21st century (National Commission on Mathematics and Science Teaching for the 21st Century, 2000). The third of the four points actually focuses on teaching, and suggests that improving teaching is a key to improving students' learning. But, consistent with Secretary Paige's report, the point is developed by recommending upgrading the quality of teachers as measured by professional and personal characteristics, not by recommending upgrading the quality of teaching methods.

One explanation for blurring the distinction between teachers and teaching is the absence of a shared, accepted knowledge base linking teaching with learning. If Secretary Paige or the Commission chaired by Senator Glenn wanted to recommend improved teaching methods, what would they say? Perhaps discussions of teaching quickly shift to teachers because not enough is known about teaching methods that support particular kinds of learning. At the same time, we believe that ignoring or blurring the conceptual distinction between teachers and teaching hinders progress in building a knowledge base. Confusion between teachers and teaching undermines research efforts to link teaching with learning.

Despite the Difficulties, Research-Based Claims Are Emerging

Given the challenges we have outlined, it now might seem surprising that empirically supported claims linking teaching with learning exist at all. However, despite the difficulties, researchers have constructed a variety of local hypotheses to guide their efforts and have used a variety of methods to investigate connections between teaching and learning. Out of this array of theoretical and empirical work, a few general patterns that link features of teaching with particular kinds of learning can be detected. This is a bold claim because it says there is sufficient empirical data, not just plausible arguments, to link some features of teaching with some types of learning.

Because the goal of this chapter is to describe empirically based connections between teaching and learning, we reviewed the literature with two primary criteria in mind. First, we wanted to select learning outcomes that were educationally significant across grade levels and specific mathematical topics. We were not concerned with detailing the kinds of teaching that promote learning of particular competencies, like solving quadratic equations or subtracting whole numbers with regrouping. Rather, we were interested in more general outcomes, like skill efficiency or problem-solving ability or conceptual understanding. If we could find links between teaching and more general outcomes, the findings would have wider applicability.

The second criterion we used to guide the review was to select only studies that described classroom teaching with some detail, that included well-described measures of students' learning, and that were designed to examine connections between the nature of the teaching and what and how well students learned. Studies that addressed well at least some of the conceptual and methodological challenges reviewed earlier were of special interest. Note that requiring studies to describe the nature of the classroom teaching eliminates a number of important areas of research on students' learning. For example, a well-established principle of learning is that the extent to which new material is learned well is influenced by what is already known (National Research Council, 1999). But few studies in mathematics education have examined these effects in the classroom by varying systematically the extent to which teaching connected with students' current knowledge or ways of thinking. Although the principle still has important implications for teaching mathematics, its empirical support does not come primarily from classroom studies of teaching and learning mathematics. As another example of excluded literatures, a rapidly increasing set of studies exists on the effects of different curricula on students' learning (National Research Council, 2004). At this point, however, most of these studies do not include detailed descriptions of classroom teaching so they are not included here. The interested reader will find the literature on the effects of curricula on students' learning reviewed by Stein, Remillard, and Smith (this volume).

In the next sections of the chapter, we present several patterns that connect mathematics teaching and learning. We begin by identifying one of the most firmly established but most general connections between teaching and learning. Commonly referred to as "opportunity to learn," this claim says that students learn best what they have the most opportunity to learn. We then increase the lens magnification slightly and identify two educationally significant learning outcomes in mathematics that have attracted considerable attention from researchers studying teaching effectiveness: skill efficiency and conceptual understanding. We identify the kinds of teaching that are empirically related to these learning outcomes.

The patterns we describe were induced by looking across many studies with an eye toward identifying convergent findings. We believe they are the patterns that enjoy the strongest empirical support at this point in our research history. Other patterns are emerging but do not yet have as much support at this point. For example, considerable work has been conducted on classroom discourse and its effects on learning (see, for example, Forman, 2003; Lampert & Cobb, 2003), but, in our judgment, the empirical links between particular kinds of discourse and particular kinds of student learning are not yet as developed as the patterns we describe. Parenthetically, we include in our review some studies that report data on classroom discourse, but we interpret the results to support a kind of teaching that includes more than just a particular kind of discourse.

Our reading of the literature indicates that the patterns we identify are sufficiently documented that they warrant action. That is, we believe they can be accepted with high enough levels of confidence that they can be used to inform policy and influence classroom practice. Saying this yet another way, they provide substantive empirically-based connections between teaching and learning that Secretary Paige or the Glenn Commission on Teaching could have used to formulate recommendations regarding improved teaching. In addition, the patterns point to fertile areas for future research. They suggest ingredients for modest theories of relationships between teaching and learning, and they map out a territory for promising empirical studies.

OPPORTUNITY TO LEARN: STILL THE KEY ENABLING CONDITION

"Opportunity to learn is widely considered the single most important predictor of student achievement" (National Research Council, 2001, p. 334). Defined by the National Research Council (2001) as "circumstances that allow students to engage in and spend time on academic tasks . . ." (p. 333), the concept of opportunity to learn has been proposed for some time as an explanation for differences in achievement among students (Floden, 2002). It has been used to account for differences in mathematics learning across a variety of contexts, including between-country differences on international comparisons. Fletcher (1971), for example, in a reanalysis of mathematics achievement data from the First International Study of Mathematics Achievement (Husen, 1967) found opportunity to learn a key explanatory construct. He summarized, "the inescapable conclusion, therefore, seems to be that reported 'achievement' is virtually synonymous with 'coverage' across countries (p 145)." Carroll's (1963) influential model for school learning included time as well as topic coverage as an important variable and promoted the concept's popularity. Researchers from inside and outside of mathematics education continue to include opportunity to learn on the short list of well-documented findings that link teaching and learning (Brophy, 1999; Grouws, 2004; National Research Council, 2001).

Opportunity to learn is not entirely a function of teaching or entirely under the control of the teacher. The curriculum the teacher is required to use, for example, surely influences students' opportunity to learn (Stein et al., this volume). But teaching, as we have defined it, plays a major role in shaping students' learning opportunities. The emphasis teachers place on different learning goals and different topics, the expectations for learning that they set, the time they allocate for particular topics, the kinds of tasks they pose, the kinds of questions they ask and responses they accept, the nature of the discussions they lead—all are part of teaching and all influence the opportunities students have to learn.

Opportunity to learn as a concept that links teaching and learning is best viewed as something more nuanced and complex than simply exposure to subject matter. Put simply, students who are exposed to a topic obviously have a better chance of learning it than students who are not. But, opportunity to learn can mean something more interesting and useful. Consider first graders exposed to a lesson on calculus. Do they have an opportunity to learn calculus? Only in an overly simplistic sense. In reality, first graders would not have a good opportunity to learn calculus because they are not sufficiently prepared to learn the material and cannot engage in the tasks and discussions that would support such learning. They might learn something by sitting through the lesson, like how to sit quietly and listen to things they do not understand. Why? Because that is what the conditions afford. It might not be the intended learning, but it is, in a real sense, what they have an opportunity to learn. So, "opportunity to learn" is not the same as "being taught." Opportunity to learn includes considerations of students' entry knowledge, the nature and purpose of the tasks and activities, the likelihood of engagement, and so on.

Vygotsky's (1978) well-known concept of the *zone of proximal development* captures the general principles behind the example of first graders being taught calculus. The zone of proximal development refers to the space within which learning can be expected, under supportive conditions, given the person's current level of functioning. It describes what is possible, given appropriate opportunities.

Opportunity to learn can be a powerful concept that, if traced carefully through to its implications, provides a useful guide to both explain the effects of particular kinds of teaching on particular kinds of learning and improve the alignment of teaching methods with learning goals. If conceptual understanding, for example, is a valued learning goal, then students will need opportunities to develop such understanding. Similar statements can be made for other learning goals, such as smooth execution of procedures, problem-solving skill, and so on. These kinds of learning do not happen automatically; students need continuing opportunities to achieve them.

Although the concept is useful, the fact that opportunity to learn still is the best link between teaching and learning signals that the research community has not advanced very far in teasing out more specific connections. In the next section, we take a small step in identifying more specific connections between features of mathematics teaching and students' learning. The two patterns we describe can be thought of as instances of opportunities to learn: opportunities to develop skill efficiency and opportunities to develop conceptual understanding.

TWO PATTERNS: TEACHING FOR SKILL EFFICIENCY AND TEACHING FOR CONCEPTUAL UNDERSTANDING

Two of the most valued learning goals in school mathematics are, and have been for some time, skill efficiency and conceptual understanding (Resnick & Ford, 1981). By *skill efficiency*, we mean the accurate, smooth, and rapid execution of mathematical procedures (see, for example, *skill learning* in Gagne, 1985). We do not include the flexible use of skills or their adaptation to fit new situations. By *conceptual understanding*, we mean mental connections among mathematical facts, procedures, and ideas (Brownell, 1935; Davis, 1984; Hiebert & Carpenter, 1992). This definition suggests that conceptual understanding grows as mental connections become richer and more widespread. As we mentioned earlier, considerable research has been conducted to identify the kinds of teaching that facilitate students' achievement of these two learning goals.

An initial and rather obvious finding is that some kinds of teaching support skill efficiency and other kinds of teaching support conceptual understanding. That different kinds of teaching facilitate different kinds of learning is a direct consequence of opportunity to learn. Different kinds of teaching provide different opportunities to learn that, in turn, yield different kinds of learning. This notion has been expressed for some time. Here is what Edward Thorndike had to say in 1912, "It is possible, by focusing attention upon immediate facility, to choose methods of teaching that are excellent for that, but defective for the more important service of arousing in a pupil the desire and power to educate himself [*sic*]" (p. 197). Gagne (1985) refined the notion by detailing different instructional techniques appropriate for different learning outcomes. And Brophy (1988) reminded educators that the work he reviewed linking teaching behaviors to students' learning mostly had used measures of routine skill learning. Other teaching features might have been identified if other learning outcomes had been used.

The story gets more complicated when one notices that neither theory nor empirical data indicate a simple correspondence between one method of teaching and skill efficiency and between another method of teaching and conceptual understanding. The best way to express current knowledge in the field is to describe some features of teaching that facilitate skill efficiency and some features of teaching that facilitate conceptual understanding, and to indicate where these features overlap or intersect.

At this point we should recognize an important historic distinction that is related to but different from the distinction in learning goals we are making here. In a classic book on learning and teaching, Ausubel (1963) proposed a 2×2 matrix with rote versus meaningful learning on one axis and discovery versus expository teaching on the other axis. Ausubel contended that these dimensions were independent. Expository teaching, said Ausubel, does not necessarily produce rote learning, and discovery teaching does not necessarily produce meaningful learning. We agree with Ausubel and raise the distinction here for two reasons. First, we want to alert readers that we are not necessarily linking skill efficiency with expository teaching. Based on our definition, skills might be executed efficiently under rote or meaningful learning conditions. Conceptual understanding, on the other hand, would require meaningful learning.

A second reason for raising Ausubel's (1963) distinction is that we do not expect the features of teaching that facilitate skill efficiency and conceptual understanding to fall neatly into categories such as "expository" or "discovery." In fact, the features of teaching we describe do not fit easily into any of the categories frequently used to describe teaching: direct instruction versus inquiry-based teaching, student-centered versus teacher-centered teaching, traditional versus reform-based teaching, and so on. Although these categories and labels have been useful for some purposes because they capture constellations of features and treat teaching as systems of interacting components, they also can be misleading because they group together features in ill-defined ways and connote different kinds of teaching to different people. For our purposes, avoiding past connotations is important. In fact, we will argue that most of these categories, distinctions, and labels are now more confusing than helpful, and further advances in research as well as clarity of policy recommendations will benefit from abandoning these labels.

But abandoning the usual labels for methods or systems of instruction in favor of identifying features of teaching that link with learning appears to violate a principle we espoused earlier—teaching is a system of interacting elements, and the effects of an individual feature are determined by the system of which it is a part. Can these two points of view be reconciled? Can researchers single out features of teaching that facilitate learning and still believe that the specific effects of individual features depend on the systems of teaching in which they operate? We think so, for two reasons. First, by looking for patterns across multiple studies, we are searching for features that replicate their effects through different implementations of similar systems and even through different systems. We suspect that the exact effects of target features vary somewhat from system to system, but we expect some of these features to emerge as sufficiently robust across systems to appear as detectable patterns. Second, we assume that when we identify specific features of teaching we actually are identifying clusters of features. The features we single out in the following sections probably operate within a supportive set of less apparent features. The regularity that can be detected in their effects on students' learning might be the result of clusters of features that function well within a range of systems of teaching. We single out specific features of teaching because the empirical data do not yet allow a more precise interpretation. But the hypothesis of systems of teaching predicts that these features are simply the most salient aspects of surrounding clusters. Perhaps the development of sharper research tools will afford more precise descriptions in the future.

Becoming Efficient in Executing Skills

No empirical studies that we are aware of set out to examine which features of teaching support skill efficiency and which support conceptual understanding. Studies have contrasted methods of teaching that emphasize skills versus understanding (e.g., Brownell & Moser, 1949; Cobb et al., 1991; Heid, 1988; Hiebert & Wearne, 1993), but no single study or set of studies aimed to optimize these two learning outcomes and sort out the features of teaching associated with each. In addition, different researchers used different measures of teaching and learning, so patterns across studies must be abstracted by interpreting designs and results and inducing consistencies, not simply by adding up the findings.

The large set of studies conducted within the process-product paradigm provides the best evidence from which to induce patterns for links between teaching and skill efficiency. The literature includes studies in mathematics and in other subjects, and, for the most part, the features of teaching that correlate with skill efficiency are similar across subjects. The most problematic aspect of interpreting this work is characterizing the nature of the dependent measures. Many studies used standardized achievement tests to assess students' learning. Although these tests often are composites of a range of items, they include a heavy skill component, require relatively quick responses, and are restricted to closed-ended, multiple choice formats. Consequently, we believe we can appropriately infer that skill efficiency, as we defined it earlier, is captured by these outcome measures.

Recall that process-product studies attempt to identify relationships between what teachers do in the classroom (the process) and what students learn as a result of this instruction (the product). To provide trustworthy data, these studies must include large sample sizes, use reliable classroom observation instruments to measure classroom processes, and assess student learning in reliable ways. We review several studies that meet these criteria and provide insights into one set of teaching strategies that is related to mathematical skill learning as measured by standardized achievement tests. The interested reader can examine a comprehensive summary of this line of research and its implications across many subject matter fields in the work of Brophy and Good (1986).

In a naturalistic study of mathematics teaching, Good and Grouws (1977) examined the teaching performance of more than 100 third- and fourth-grade teachers over a 2-year period. An initial finding was that the impact of many teachers on students' mathematics learning, as measured on a standardized achievement test, was not very stable over time. This finding suggests that their teaching methods might have changed from year to year, perhaps due to the changing composition of their classes. The authors were able, however, to identify a group of teachers who consistently facilitated better than statistically predicted achievement results over 3 consecutive school years[2] and teachers who were less successful than would have been expected during the same time period. Results from the study indicated that teaching effectiveness was associated with the following behavioral clusters: whole-class instruction[3]; general clarity of instruction; a nonevaluative, task-focused en-

[2] The teachers were identified by using achievement data over a two year period and their success was confirmed for the third year during the course of an observational study.

[3] Whole class instruction was the organizational mode in the high classrooms (and low classrooms) with teachers employing instruction via groups in the middle of the distribution.

vironment; higher achievement expectations (more homework, fast pace); student initiated behavior; and classrooms relatively free of behavioral problems. The faster pace finding was also identified in a study of fifth-grade classrooms conducted by McDonald (1976) and singled out by Leinhardt (1986) as a characteristic of teaching associated with standardized achievement gains. Other common findings in the literature are illustrated in the next process-product study we examine.

Using the process-product paradigm, Evertson, Anderson, Anderson, and Brophy (1980) studied junior high school teachers of mathematics (N = 29) and English (N = 39). They used both low-inference and high-inference observational coding schemes and measured student learning using the California Achievement Test (CAT). Effective mathematics teachers asked more questions than did less effective teachers with most of the questions lower-order product questions. The authors summarized that "the most effective [mathematics] teachers were active, well organized, and strongly academically oriented" (p. 58). Further, as in the previous described study, the effective teachers tended to emphasize whole-class instruction with some time devoted to seatwork. In general, they actively managed their class efficiently. Although the terminology used in this study was different from that in the Good and Grouws (1977) study, some obvious parallels exist in their findings.

A series of studies conducted by researchers at the Far West Regional Laboratory for Educational Research and Development during the 1970s focused on linking teacher behavior to student achievement in mathematics and reading. This research became known as the Beginning Teacher Evaluation Study (BTES), although in practice the studies involved teachers with at least 3 years of experience and concentrated on research rather than evaluation. Interestingly, none of the teacher behaviors measured was a significant predictor of student learning in both mathematics and reading. This finding lends credence to the position that some important teaching actions are content specific.

The results from the BTES about mathematics teaching and learning tend to agree with the previously cited studies but add some interesting findings when data are examined at the individual student level. The studies found a positive relationship between student achievement on a standardized achievement test and high success rates when engaged in mathematical work during classroom lessons (correct answers except for occasional errors). Conversely, there was a negative relationship when students were engaged in activities with low success rates. The researchers considered this an important enough finding to refine the concept of opportunity to learn to mean "the amount of time a student spends engaged in an academic task [related to the intended learning] that s/he can perform with high success" (Fisher et al., 1980, p. 8). Interestingly, the attention given to high success rates in these studies provided the rationale for Good and Grouws emphasizing the importance of students being well prepared for seatwork before working independently. An overview of the studies in the BTES can be found in Powell (1980) with a detailed account provided in the five volume final report by McDonald and Elias (1976).

The features of teaching that emerge from the studies just reviewed can be summarized as follows: teaching that facilitates skill efficiency is rapidly paced, includes teacher modeling with many teacher-directed product-type questions, and displays a smooth transition from demonstration to substantial amounts of error free practice. Noteworthy in this set of features is the central role played by the teacher in organizing, pacing, and presenting information to meet well-defined learning goals.

The features of teaching identified above facilitate students' skill efficiency as measured by standardized tests. What effects these features have on students' conceptual development is unclear. Because conceptual understanding is of great interest in mathematics education and infuses the recently conceived goal of mathematical proficiency (National Research Council, 2001), we devote considerable space in the next section to examining links between teaching and students' conceptual understanding.

Developing Conceptual Understanding

We defined conceptual understanding earlier as the mental connections among mathematical facts, procedures, and ideas (Brownell, 1935; Davis, 1984; Hiebert & Carpenter, 1992). This simple definition works well for our purposes but is not without controversy. Some researchers argue that this definition places too much emphasis on the internal or mental representations of individuals and that social/cultural factors should be considered (Lave, 1988; Wenger, 1998). From this sociocultural view, understanding can be seen as an activity of participating in communities of people who are becoming competent in the practices of the trade. In our case, this would mean participating in a community of people who are becoming adept at doing and making sense of mathematics as well as coming to value such activity. We believe that both the mental connections and the cultural participation views of understanding can be useful (Cobb, 1994; Sfard, 1998) and that the claims we make in the following sections are reasonable under either definition. However, because the research examples we review were conducted within the then-prevailing mental connections view, this definition should be kept in mind when reading the next sections.

Teaching That Promotes Conceptual Development: Two Key Features

Educators always have been interested in promoting conceptual understanding, at least in their rhetoric, but there has been a recent explosion of interest in the topic. On the day we drafted this paragraph, amazon.com found 90,448 book entries for the terms: teaching for understanding. Among the best-selling titles were *Teaching for Musical Understanding*, *Strategies That Work: Teaching Comprehension to Enhance Understanding*, and *Shakespeare Without Fear: Teaching for Understanding*. Two well-known books that contain material relevant for this chapter are *Teaching for Understanding* edited by Cohen, McLaughlin, and Talbert (1993) and *Mathematics Classrooms That Promote Understanding* edited by Fennema and Romberg (1999). Clearly, the topic does not lack for analyses and recommendations.

The research base that links features of teaching with students' understanding is much thinner than the current attention might suggest. Looking across the landscape of research connecting teaching and learning, we agree with Brophy's (1997) observation that the research conducted provides less insight into how students construct understanding during instruction than in how they develop skill efficiency. This is due, in part, to the challenge of creating measures that assess conceptual understanding and the fact that much past work has used standardized tests and other similar measures that are more skill oriented. Nevertheless, we can identify two features that emerge from the literature as keys to promoting conceptual understanding. We view our presentations of both features as plausible conclusions drawn from past research *and* as descriptions of domains in which future research could be especially rich and useful.

Feature 1: Teachers and Students Attend Explicitly to Concepts

A clear pattern across a range of empirical studies is that students can acquire conceptual understandings of mathematics *if* teaching attends explicitly to concepts—to connections among mathematical facts, procedures, and ideas (Gamoran, 2001; Hiebert, 2003; National Research Council, 2001). By *attending to concepts* we mean treating mathematical connections in an explicit and public way. Brophy (1999) described such teaching as infused with coherent, structured, and connected discussions of the key ideas of mathematics. This could include discussing the mathematical meaning underlying procedures, asking questions about how different solution strategies are similar to and different from each other, considering the ways in which mathematical problems build on each other or are special (or general) cases of each other, attending to the relationships among mathematical ideas, and reminding students about the main point of the lesson and how this point fits within the current sequence of lessons and ideas.

To elaborate our claim and to illustrate the findings that support it, we will review several studies that typify those upon which the claim is based. More extensive reviews of this literature can be found in *Adding It Up* (Chapters 6, 7, and 8, National Research Council, 2001), in *A Research Companion to Principles and Standards for School Mathematics* (Kilpatrick, Martin, & Schifter, 2003), in *Mathematics Classrooms That Promote Understanding* (Fennema & Romberg, 1999), and in a chapter by Hiebert and Carpenter (1992) in a previous handbook for teaching and learning mathematics.

In many ways, the claim that students acquire conceptual understanding of mathematics if teaching attends explicitly to mathematical concepts is a restatement of the general observation that students learn what they have the best opportunity to learn. At a surface level, the claim seems rather obvious and, perhaps, not even very interesting. But it becomes more interesting when one discovers that the claim is supported across a wide range of research designs and holds true across different instructional treatments or systems. The studies we review, and the larger literature from which these studies are drawn, display a remarkable variation. Variability across studies can be described in the following ways:

- The effects appear in studies of varying designs—from tightly controlled studies of shorter duration to more loosely designed studies of longer duration;
- Both teacher-centered and student-centered teaching systems or methods (using the old labels) that have attended to conceptual development have shown higher levels of students' conceptual understanding than similar methods that have not attended to conceptual development as directly;
- The ways in which concepts are developed in classrooms can vary—from teachers actively directing classroom activity to teachers taking

> less active roles, from methods of teaching that highlight special tasks and/or materials to those that highlight special forms of classroom discourse to those that highlight student invention of solution strategies.

We believe the evidence does not justify a single or "best" method of instruction to facilitate conceptual understanding. But we believe the data do support a feature of instruction that might be part of many methods: explicit attention to conceptual development of the mathematics.

To support and clarify the claim, we chose the following first set of studies to review because they focused on the same mathematical topic—multidigit addition and subtraction in the primary grades—but implemented quite different treatments using different research designs. Fuson and Briars (1990) investigated the effects of a particular conceptual instruction approach on students' acquisition of the standard addition and subtraction algorithms. Specially designed instruction on addition was given to 169 first and second graders, and 75 of the second graders also received special instruction on subtraction. Instruction included careful demonstrations of the algorithm using base-10 blocks, with special attention to the pattern of the regrouping as the numbers increased to three and four digits. Connections were explicitly drawn between the physical and written representations. Some teachers used whole-class formats with teacher demonstrations and class discussions, and other teachers asked students to work in small groups and complete some problems on their own. In all classrooms, the focus was on understanding the standard algorithm, not inventing new strategies. Before instruction, and then after 3–6 weeks of instruction on place value and addition, plus an additional 2–4 weeks for those receiving instruction on subtraction, students were assessed using a traditional written skills test, a specially designed written test on underlying concepts, and individual interviews. As might be expected for students this young, pretest performance was quite low. For example, only 9 of the 169 students correctly regrouped when completing a multidigit addition task. In contrast, performance was uniformly high on the posttest. For example, 160 of the 169 students correctly regrouped on a similar posttest item. Even more striking were the ceiling levels of performance on conceptually based items, including those on the interviews asking students to explain why and how regrouping works in the multidigit algorithm. Given the high levels of performance across the full sample, the particular ways in which the teachers implemented the instruction (e.g., whole class demonstrations versus small group discussions) apparently did not substantially affect the outcomes. The key seemed to be the explicit attention to the conceptual underpinnings of the algorithm through connecting written and physical representations.

The absence of a control group in the Fuson and Briars (1990) study leaves many questions unanswered. These are answered, in part, as the authors point out, by the fact that the level of performance by the first graders and the (lower achieving) second graders in their study was higher than that of many third and fourth graders on similar items reported on the National Assessment of Educational Progress. This comparatively high performance supports claims about the effectiveness of the instruction. In addition, the findings were replicated by the authors (reported in the same article), with only slightly lower levels of performance, with 2723 second graders in the Pittsburgh City Schools. Clearly, the effects of this instruction were not an artifact of a special sample. In sum, the findings can be viewed as an existence proof that attending to conceptual underpinnings—to connections among mathematical ideas and representations—can facilitate students' conceptual understanding. The results also show strong skill learning, a finding that we will explore in a later section of the chapter.

A second study, conducted 40 years earlier by Brownell and Moser (1949), employed control groups and fills in another part of the picture on the effects of instruction that attends explicitly to conceptual development. Brownell and Moser studied the effects of teaching multidigit subtraction algorithms "meaningfully" versus "mechanically." About 1400 third graders in 41 schools were involved in the study. Two different algorithms, equal additions (adding equal amounts to the minuend and subtrahend to keep the difference the same) and decomposition (the current standard U.S. algorithm of regrouping), were crossed with two teaching approaches (meaningful and mechanical) to create four conditions. Teachers in the 41 schools who volunteered to participate indicated which of the approaches they preferred. Classrooms were assigned to conditions by taking into account teachers' preferences and by trying to equalize students' entry abilities across conditions. Each lesson for each condition during the 15-day experimental period was prescribed, and all lessons were quite teacher-directed. The meaningful or conceptual approach included the use of physical objects (bundles of sticks) to illustrate grouping by tens, writing numerals in expanded notation and connecting written and physical representations, and delaying the step-by-step algorithm until after examining the meaning of the number manipulations. The mechanical approach began with a demonstration of the step-by-step algorithm and did not include a rationale for the procedure. The time saved by this approach was used for practicing the procedures. Students' learning was assessed immediately after the instruction and 6 weeks later using written computation tests and individual interviews. Items measured speed, accuracy, explanations of procedures, and transfer to new kinds of subtraction problems. The results are complicated, but, in general, the following conclusions can be drawn. On the immediate measures of speed and accuracy, the mechanical approach produced higher scores. On measures of retention and transfer, the meaningful approach produced higher scores. The meaningful approach worked better for the decomposition algorithm than the equal additions algorithm and for students who had not already memorized a subtraction algorithm before they began instruction. In general, if good explanations, retention, and transfer are valued learning goals (Brownell, 1948), then the meaningful approach showed the most promise, regardless of which algorithm was taught.

Critics might argue that the results reported by Brownell and Moser (1949) are not conclusive because the mechanical approach was highly constrained and the large sample of teachers, without classroom observations to suggest otherwise, could have implemented the approaches in quite different ways. The criticisms are appropriate and should be considered when interpreting the results. Still, we can reasonably conclude that explicit attention to conceptual underpinnings resulted in higher levels of conceptual understanding. This reinforces the conclusion drawn from Fuson and Briars (1990) using a much different research design and separated in time by more than 40 years.

A third study on multidigit addition and subtraction fills in another part of the picture. Hiebert and Wearne (1993) examined the effects of a mixed teacher- and student-centered form of conceptually-based instruction on conceptual understanding (and procedural skill). The 147 second graders in the six classrooms of one school were assigned to one of two treatment conditions for the 12 weeks normally devoted to multidigit addition and subtraction by the teachers. Conceptually-based instruction focused on building a foundation of place value understanding by connecting written and physical (base-10 blocks) representations of numbers and then asked students to invent their own procedures for solving addition and subtraction problems. Teachers demonstrated the standard algorithms if they were not presented by students. Traditional instruction followed the textbook as ordinarily used by the teachers. The focus in these classrooms was on demonstrating and practicing the standard algorithms. Four classrooms (one high achieving and three average-low achieving) received the conceptually based instruction taught by specially trained teachers and two classrooms (one high achieving and one average-low achieving) received the traditional instruction taught by the classroom teachers. Weekly observations and audiotapes recorded the tasks used during the lessons and the talk by teachers and students in all the classrooms. Instruction, as implemented, showed differences between the number and kind of tasks presented and the nature of the classroom discourse. In general, students receiving the conceptually based instruction spent a greater percentage of time working on the rationale for procedures and examining the legitimacy of invented procedures. Students receiving the traditional instruction spent a greater percentage of time practicing taught procedures. On the basis of beginning-of-the-year and end-of-the-year tests, students in the conceptually based classrooms gained more on skill items and, especially, on conceptual items such as explaining the rationale for computation procedures and transferring them to new kinds of problems. By the end of the year, the performance profile of the lower achieving classroom that received the conceptually based instruction looked much like the higher achieving traditionally instructed classroom.

The use of specially trained teachers for the conceptually based instruction and the relatively small sample size in Hiebert and Wearne (1993) lead to less-than-conclusive results but for different reasons than in Fuson and Briars (1990) or in Brownell and Moser (1949). Each study contains some design deficiencies, but the deficiencies lie in different places. That is why we avoid drawing conclusions about the links between the specific kinds of instruction on multidigit addition and subtraction described in each study and the acquisition of specific skills and understandings. But we do see a pattern across these studies, and others like them, in the general feature of attending to conceptual development. We conclude that when teaching attends explicitly and directly to the important conceptual issues students are more likely to develop important conceptual understandings. This conclusion is reinforced further by considering a few additional and very different kinds of studies.

The evidentiary base supporting conceptual teaching is not restricted to research in the computational skills domain. In a summary report of three separate experimental studies of mathematics learning defined broadly in classrooms at Grades 4, 6, and 8, respectively, Good, Grouws, and Ebmeier (1983) found substantial support for an instructional program that was teacher directed, tightly controlled, and emphasized attention to conceptual development. To define (conceptual) development, the authors drew on the early work of Brownell (1947) on meaningful learning and the subsequent recapitulation of this work, termed *relational understanding*, by Skemp (1976). They summarized their use of the term *development* as "that part of a lesson devoted to increasing comprehension of skills, concepts, and other facets of the mathematics curriculum" (p. 203).

Common to the set of studies reported by Good et al. (1983), and quite different from the preceding studies reviewed in this section, was the use of a treatment that asked teachers to follow a prescribed, highly structured daily teaching plan that employed whole class instruction and included a brief introductory segment (attend to homework, short review, mental computation practice), a development component (about one half the remaining class period), seatwork, lesson closure, and assignment of homework. In the studies, intact classrooms were randomly assigned to treatment and control conditions. Control group teachers were asked to teach using their usual teaching methods.

A number of unique features of these studies are worth noting when interpreting the findings. First, total test scores and subtest scores of the mathematics portions of standardized achievement tests were the main dependent measures used to determine student learning, although as will be mentioned later, supplementary content measures were used in some studies. Second, analysis of variance on residualized gain scores were used to examine the pre-post test differences on each measure. Third, class means were used as the unit of analysis in the statistical tests employed. Finally, interpretation of results took account of treatment fidelity using data from various combinations of classroom observations, teacher diaries, and self-report questionnaires.

In the fourth-grade study of 40 classroom teachers, the treatment teachers exhibited substantially more of each of the requested teaching features than did the control teachers. Furthermore, the scores of the students of the treatment teachers showed significant improvement from pretest to posttest moving from the 27th percentile to the 57th percentile on national norms, significantly higher than the student scores from the control group (p = .003) (Good & Grouws, 1979). On a specially designed content test developed by an outside consultant to measure specific content in the curriculum including skills, concepts, and verbal problem solving, the treatment group again outperformed the control group, but the statistical difference was more modest (p = .10) with only small differences between the treatment group and the control group on the verbal problem-solving items.

In the sixth-grade study involving 36 teachers and in the eighth-grade junior-high study involving 19 teachers (41 classes), the treatment was modified to give regular, systematic attention to verbal problem solving. In both studies, the treatment students outperformed the control students on the standardized achievement test and on a research-constructed measure of verbal problem solving. The differences on both measures, however, were not as large as in the fourth-grade study. The authors hypothesized that the smaller effects in the sixth-grade study might have been due to smaller differences between teaching methods in the treatment and control classrooms than in the original study. They attributed the lack of larger teaching differences to the possibility of sixth-grade teachers visiting with and learning from the fourth-grade treatment teachers who participated in the study the previous school year. The somewhat diminished fidelity of Grade 8 teachers to the teaching expectations requested in the treatment might be due, according to the authors, to a shorter training period than in previous studies and to a possible treatment request overload—the teachers were asked to implement all the teaching requests in the previous study plus the requests associated with increased attention to verbal problem solving. In spite of less fidelity to treatment than in the previous studies, the researchers did find a statistically significant difference between treatment and control classrooms on the problem-solving measure at the sixth-grade level when 7 poorly implementing teachers were removed from the analysis.

The researchers are quick to point out that they "are *not* suggesting that the instructional program used in the study is the best approach to take for facilitating the mathematics achievement of students" (Good et al.,1983, p. 91), but rather that the ideas within the program including more attention to conceptual development, systematic attention to review, and utilization of appropriate homework assignments might be of considerable value to teachers who use a whole-class organizational pattern for teaching mathematics. The important message from these studies for the current discussion is the recurring link between explicit attention to the development of mathematical connections among ideas, facts, and procedures and the increased conceptual learning by students. The studies not only support this contention but also seem to show that such instruction facilitates the development of important mathematical skills beyond those associated with computational skills, a point to which we will return in a later section.

In contrast to the rather tightly controlled and precisely targeted designs of the studies reviewed above, Fawcett (1938) and Boaler (1998) reported multi-year studies on the cumulative effects of instruction at the secondary level. Fawcett described a 2-year experimental course in high-school geometry for which instruction was redesigned to begin with the reasoning of students and build toward conventional deductive forms of reasoning through, in part, analyzing the logic of real-life arguments. Conceptual aspects of this instruction are apparent in the explicit attention to connections among mathematical objects—definitions, arguments, proofs, claims, and so on. Performances of the 25 students in this course were compared to those of a similar number of students in a more traditional (2-year) course in geometry. On conventional measures of geometric content, on retention of content, and on transfer measures involving the analysis of reasoning in nonmathematical situations, students in the experimental group outperformed students in the control group. Although many design features undermine the comparative claims (e.g., the absence of information about the instruction provided in the more traditional course), a fact acknowledged explicitly by Fawcett (1938), it does seem clear that students in the experimental group engaged with important mathematical ideas and developed important understandings.

Boaler's (1998) study, 50 years later, addressed some of the design deficiencies of Fawcett's (1938) study and reached similar conclusions. Boaler followed a cohort of students at two different secondary schools for 3 years. The schools were chosen because they claimed to be implementing different forms of instruction. After observing regularly lessons in both schools, Boaler described the instruction in one school as an efficient and well-ordered textbook-based approach to procedural competence. Teachers demonstrated procedures and students practiced them. Instruction in the other school was directed toward more conceptual learning goals and revolved around open-ended projects. Students often worked cooperatively, sharing and explaining solution strategies. Entry measures showed that the approximately 200 students in the first school and the approximately 100 students in the second school performed similarly when they began. But, after 3 years, students in the second school, emphasizing conceptual problems and discussions, performed better than their peers on both specially designed conceptual measures and on standardized tests.

In summary, the findings from these studies typify those found in much of the literature that assesses the effects of instruction that explicitly attends to conceptual development of mathematics. Students receiving such instruction develop conceptual understandings, and do so to a greater extent than students receiving instruction with less conceptual content. As noted earlier, this claim might not be surprising because it is a straightforward instantiation of the principle of opportunity to learn. It becomes more interesting, however, when one notes that conceptual development of the mathematics can take many pedagogical forms. Concepts can be developed through teacher-centered and highly structured formats as in some of Fuson and Briar's (1990) classrooms and in Good et al.'s (1983) classrooms or through student-centered and less structured formats as in Fawcett's (1938) classroom. The claim becomes even more interesting and important when making two additional observations that will be developed later in the chapter. First, as we have noted, the findings from several of the studies we reviewed suggest that instruction emphasizing conceptual development facilitated skill learning as well as conceptual understanding. Second, although instructional attention to conceptual development might seem like an obvious feature of "good" teaching and might be expected in most classrooms, it is a feature consistently absent in descriptions of U.S. mathematics classrooms. Before exploring these two issues, we present the second feature of instruction that is associated consistently with increased conceptual understanding.

Feature 2: Students Struggle with Important Mathematics

Our interpretation of the literature on teaching for conceptual understanding points to a second feature of teaching that consistently facilitates students' conceptual understanding: the engagement of students in struggling or wrestling with important mathematical ideas. Unlike the first feature, this second feature might not be obvious to readers, so we first clarify what we mean by *struggle*, then elaborate the theoretical connection between struggling with and understanding mathematics, and finally review a small sample of empirical studies from this perspective.

We use the word *struggle* to mean that students expend effort to make sense of mathematics, to figure something out that is not immediately apparent. We do *not* use *struggle* to mean needless frustration or extreme levels of challenge created by nonsensical or overly difficult problems. We do not mean the feelings of despair that some students can experience when little of the material makes sense. The struggle we have in mind comes from solving problems that are within reach and grappling with key mathematical ideas that are comprehendible but not yet well formed (Hiebert et al., 1996). By struggling with important mathematics we mean the opposite of simply being presented information to be memorized or being asked only to practice what has been demonstrated.

Dewey (1910, 1926, 1929) devoted a good deal of attention to elaborating the processes and consequences of this kind of struggle. For Dewey, the process of struggle was essential for constructing deep understandings. The process begins, said Dewey (1910) with "some perplexity, confusion, or doubt (p. 12)." It continues as students try to fit things together to make sense of them, to work out methods for resolving the dilemma. Traditional school instruction, noted Dewey (1926), is plagued by a press for quick answers. This short-circuits opportunities to think about the conceptual aspects of the problem. Deep knowledge of the subject, said Dewey (1929), "is the fruit of the undertakings that transform a problematic situation into a resolved one" (p. 242–243).

Other cognitive theorists have developed similar arguments that connect the process of struggling to make sense of the subject with deeper understandings of the subject. Festinger's (1957) theory of cognitive dissonance developed the notion of cognitive perplexity as a central impetus for cognitive growth. In a more targeted way, Hatano (1988) identified cognitive incongruity as the critical trigger for facilitating the development of reasoning skills that display conceptual understanding. For Hatano, cognitive incongruity places the student in the situation described by Dewey as filled with perplexity, confusion, or doubt. Applying a similar theoretical construct to school classrooms, Brownell and Sims (1946) contrasted the repetitiveness of most school tasks with the processes they argued are essential for developing understanding—students must feel a need to resolve a problematic situation and then be allowed "to muddle through" (p. 40). What kinds of tasks or problems pose appropriate problematic situations? Vygotsky's (1978) zone of proximal development provides a guide for answering this question because it can be interpreted as the space within which a student's struggle is likely to be productive. That is, novel problems near the boundaries of a student's zone of proximal development are problems that are within reach but that present enough challenge, so there is something new to figure out.

Mathematicians and others who are concerned about how students interact with the discipline of mathematics have alluded to struggling with key mathematical ideas as a natural part of doing mathematics. In his classic book on problem-solving heuristics, Polya (1957) echoed themes of Dewey (1910) by opening the preface to the book's first printing with the following:

> A great discovery solves a great problem but there is a grain of discovery in the solution of any problem. Your problem may be modest; but if it challenges your curiosity and brings into play your inventive faculties, and if you solve it by your own means, you may experience the tension and enjoy the triumph of discovery (p. v).

Polya continued by noting that experimentation and struggle are an integral part of developing the discipline of mathematics and suggested that students should experience this side of mathematics as well as the systematic, deductive side. Brown (1993) explicitly brought together Dewey's and Polya's views around the notion of struggling with mathematics, a phenomenon Brown called "confusion." He suggested that struggling with real problems is quite different than the tidy curriculum students usually receive, and he argued that appropriate struggle (or confusion) reveals to students a key aspect of mathematics. "We seem incapable of appreciating what it means to be confused, what virtues may follow from it and what kind of pedagogy may make use of it" (Brown, 1993, p. 108).

The logic that connects struggle with understanding, the learning goal of interest in this section, can be expressed using either of the definitions of *understanding* presented earlier. If understanding is defined as the mental connections among mathematical facts, ideas, and procedures, then struggling is viewed as a process that reconfigures these things. Relationships among facts, ideas, and procedures are re-formed when new information cannot easily be assimilated or when the old relationships are found to be inadequate to make sense of a new problem (Piaget, 1960; Skemp, 1971). Struggle results in restructuring one's mental connections in more powerful ways. If understanding is defined as participating in a community of people who are becoming adept at doing and making sense of mathematics, then struggling is vital because it can be an essential aspect of personal meaning-making within the community (Handa, 2003). It is central to what mathematicians do (Brown, 1993; Polya, 1957). In summary, struggling with important mathematics is implicated in both definitions of understanding by identifying the common processes hypothesized to develop understanding.

Although little, if any, research has tried to isolate and test the effects of struggle on the development of students' conceptual understandings, many findings suggest that some form of struggle is a key ingredient in students' conceptual learning. In a provocative review of cognitive training studies (outside of mathematics education), Robert Bjork (1994) pointed out a curious feature in many of the findings: When trainees experienced difficulties and could not smoothly master the targeted skills, they developed deeper or more useful competencies in the long run. The difficulties were introduced either intentionally or accidentally into the training procedure and took different forms—changing the context of problems, withholding feedback, introducing interfering distractions, and so on. The key condition was that trainees faced some obstacles along the way. In these cases, the trainees generalized their learning more appropriately and transferred it more effectively to novel tasks. Apparently the trainees had to think harder about what they were learning. Bjork (1994) concluded:

> Whatever the exact mixture of manipulations that might turn out to be optimal [for learning], one general characteristic of that mixture seems clear: It would introduce many more difficulties and challenges for the learner than are present in typical training routines (p. 189).

This conclusion is not surprising, said Bjork, because it simply affirms the observation, now widely accepted in cognitive psychology, that learners construct their own interpretations and memories of activities rather than simply absorbing them from the environment. When students struggle (within reason), they must work more actively and effortfully to make sense of the situation, which, in turn, leads them to construct interpretations more connected to what they already know and/or to reexamine and restructure what they already know. This yields content and skills learned more deeply.

No series of classroom studies parallels those of the laboratory studies in the training literature reviewed by Bjork (1994), but, when viewed through the lens of students' struggle, the findings from a number of classroom studies are suggestive. Again, we review only a few studies here to illustrate the nature of these findings. First, in a study outside of mathematics, Capon and Kuhn (2004) found that students in a graduate MBA course who learned a new business concept through attempting to solve problems rather than listening during a lecture and discussion could apply what they learned to more effectively explain a related concept. The authors accounted for the results by arguing that working to solve problems promoted more sense making and integrating of information, in other words more conceptual understanding.

We now review several studies in mathematics education that support the claim that some features of teaching can facilitate students' opportunities to struggle, in the sense that we have defined it, and when this happens students develop greater conceptual understanding. Although Carpenter, Fennema, Peterson, Chiang, and Loef (1989) did not interpret their data from this perspective, they reported findings that are consistent with our claim. The authors compared the teaching and learning in 40 first-grade classrooms. Half of the teachers from these classrooms were randomly selected to receive a 4-week summer course (called Cognitively Guided Instruction, or CGI) that focused on young children's knowledge and skills of simple addition and subtraction. Classroom observers recorded the kinds of teaching that occurred for addition and subtraction the following year, and the researchers administered a variety of student achievement measures. Teachers who had received the 4-week course spent a higher percentage of instruction time on word problems (often thought to be more challenging for students) whereas the control teachers spent a higher percentage of instructional time on number facts problems and on review. In addition, the CGI teachers posed more arithmetic problems to the whole class, listened more often to the solution procedures presented by students, and gave feedback less often on whether students' answers were correct. All these features of teaching are consistent with the hypothesis that students in the classrooms of CGI teachers had more opportunities to wrestle with the mathematics involved in adding and subtracting whole numbers. On many of the achievement measures, there were no significant differences between students in the two groups. However, students in the classes taught by CGI teachers performed better on the "complex addition and subtraction problems," problems requiring more than routine knowledge of arithmetic. Interestingly, students in these same classrooms also demonstrated significantly higher recall of number facts during individual interviews than students in the control classrooms. Although only suggestive, the results reported by Carpenter and colleagues (1989) are consistent with our primary claim in this section.

A study that addresses more directly the claim that when teaching promotes constructive struggle with mathematics students' understanding increases was conducted as part of the QUASAR Project (Silver & Stein, 1996; Stein & Lane, 1996). In particular, the results of the study suggest that mathematical tasks that make higher cognitive demands on students facilitate greater conceptual development (Stein, Grover, & Henningsen, 1996; Stein & Lane, 1996). The mathematical tasks presented to middle-school students were examined for the cognitive effort, or "thinking processes," the tasks were likely to elicit. The most challenging tasks were presumed to require reasoning mathematically (problem solving, pattern finding, conjecturing) and connecting procedures with concepts. The least challenging tasks were presumed to require executing procedures without considering connections to other ideas, memorizing facts and rules, and exploring unimportant mathematics or exploring in a haphazard, unpurposeful way. The level of challenge of a sample of 144 tasks presented in 3 classrooms in each of 4 middle schools over 3 years was correlated with the achievement gains by sixth, seventh, and eighth graders tested at the beginning and end of each school year. Achievement was measured using a specially designed pool of 36 open-ended mathematics problems that allowed tracing changes in students' performance over several years. Results are quite complicated but clearly indicate that students attending schools in which teachers presented *and* faithfully implemented more challenging problems were more likely to develop increased conceptual understanding of the mathematics.

The study by Hiebert and Wearne (1993), reviewed earlier, supplements the findings of the QUASAR studies because teachers in the classrooms that facilitated greater understanding asked a higher percentage of questions that required explanations and analyses, and students in these classrooms responded with longer utterances. These forms of discourse require more intellectual work from students and can reveal students who are struggling to make sense of mathematics and to express their emerging understandings.

That students can make sense of classroom discourse that displays genuine cognitive and mathematical struggle was documented nicely by Inagaki, Hatano, and Morita (1998). Fourth and fifth graders in 11 classrooms were observed during discussions of alternative methods for adding fractions with unlike denominators. Although the discussions were filled with confusions and incorrect conjectures along with correct analyses, most students made sense of the discussions and improved their understanding as demonstrated through both their verbal statements and written work.

Case studies of classrooms in which teachers engage students in struggling with important mathematics add support to the claim that struggle is a key enabler for conceptual understanding. Detailed descriptions of classrooms in which authors reflect on their own experiences as teachers are especially compelling (Ball, 1993; Fawcett, 1938; Heaton, 2000; Lampert, 2001; Schoenfeld, 1985). Classrooms in which this kind of struggle is noticeably absent and, in turn, students do not improve their understanding (Schoenfeld, 1988) complement the positive cases.

A number of questions regarding classrooms in which students are engaged in constructive struggle with mathematics remain unanswered. In contrast to the first feature of explicit attention to concepts, whether struggle can be incorporated into a wide range of pedagogical forms is unclear. Struggle is usually associated with student-centered or student inquiry approaches. But we can imagine teacher-centered approaches that provide targeted and highly structured activities during which students are asked to solve challenging problems and work through challenging ideas. In fact, it seems plausible that students' struggle should be sufficiently bounded and directed so that it centers on the important mathematical ideas. This requires some level of teacher guidance. There are few principles, however, from theory or data, that speak to the degree of structure needed to facilitate (or undermine) the productive effects of students' struggle. Parenthetically, the possibility that appropriate struggle can be built into teacher-centered approaches is one reason why we think the old labels of student-centered versus teacher-centered instruction can be so misleading.

A striking contrast should be apparent between teaching that encourages students to struggle and teaching for skill efficiency, described earlier, that ensures high success rates. Although these are not completely contradictory features, they would likely be found in quite different systems of teaching. On the one hand, this is consistent with what one might expect—different features and clusters of features are effective for promoting skill efficiency than are effective for conceptual understanding. But what explanation can account for the fact that skill learning was relatively high in classrooms where features of teaching associated with conceptual development were being implemented? We turn next to consider this question.

Teaching Features that Promote Conceptual Understanding also Promote Skill Fluency

We noted earlier that no simple correspondence exists between features of teaching and learning outcomes. One place in which the complex nature of teaching and learning becomes apparent is in the effects of conceptually oriented teaching on skill learning. Many of the reports on the conceptual development of students also indicate that their skills increased at a level equal to or greater than students in the control groups (Boaler, 1998; Fawcett, 1938; Fuson & Briars, 1990; Good et al., 1983; Hiebert & Wearne, 1993; Stein & Lane, 1996). Apparently, it is not the case that only one set of teaching features facilitates skill learning and another set facilitates conceptual learning. In this case, two quite different kinds of features both seem to promote skill learning.

One way to explain this observation is that the *nature* of skill learning might be somewhat different under the two instructional conditions. Under one condition, instruction is quickly paced, teachers ask short-answer targeted questions, and students complete relatively large numbers of problems during the lesson with high success rates. Under the other condition, instruction is more slowly paced, teachers ask questions that require longer responses, and students complete relatively few problems per lesson. Using Wittrock's (1986) argument, the cognitive mechanisms that students are likely to use under these different conditions would be different, leading to different skill competencies. In fact, Bjork's (1994) review seemed to show exactly that: Trainees who mastered skills under conditions that stimulated more conceptual work acquired different competencies—they were better able to adapt their skills to solve new kinds of tasks. Following the authors of *Adding It Up* (National Research Council, 2001), we use the word *fluent* (rather than *efficient*) to describe this kind of competence.

The foregoing analysis produces the following conjecture. The two features we identified for developing conceptual understanding are promising candidates for facilitating a more ambitious learning goal—mathematical proficiency (National Research Council, 2001). *Mathematical proficiency* is the proficiency achieved through integrating five kinds of competencies, including conceptual understanding and procedural fluency.[4] Because mathematical proficiency has not been defined long enough to serve as the outcome measure for empirical studies, no unambiguous data are available to test the relationships between teaching and students' mathematical proficiency. But the findings of improved conceptual understanding *and* procedural fluency related to the two features of teaching we described above make plausible the conjecture that these two features can facilitate the development of mathematical proficiency.

Summary

We believe that two features of classroom mathematics teaching facilitate students' conceptual development (and perhaps mathematical proficiency)—explicit attention to connections among ideas, facts, and procedures, and engagement of students in struggling with important mathematics. We believe both features are justified by not only the theoretical arguments but also the empirical data. We have not identified all of the teaching actions that might call students' attention to relevant connections or engage them in productive struggle, but we have described a sample of strategies (e.g., posing problems that require making connections and then working out these problems in ways that make the connections visible for students). The centrality of the teacher was noted when reviewing the teaching features that facilitate the acquisition of skill efficiency. On the basis of the current literature, systems of teaching that facilitate the development of understanding seem to operate with the teacher playing both more and less active roles.

Our reading of the literature suggests that other features of teaching that sometimes are associated with conceptual development (e.g., use of concrete materials, asking higher order questions) are too specific and tied too closely to particular classroom conditions to support claims that apply across classrooms. That is, not enough evidence exists to conclude that these more specific features, which work well under certain conditions or within particular systems of teaching, will work well under other conditions or in other systems. The two features we identified are more general and seem to operate effectively across a range of contexts and teaching systems.

We have not commented on whether both teaching features—attending to connections and struggling with important mathematics—are necessary for promoting conceptual understanding, whether either one is sufficient, whether there are interactions between the two, and so on. We also have not speculated about how these features might function as part of clusters of features. Perhaps, for example, each of these two features is part of a cluster that travels, as a cluster, across different systems of teaching. Increased conceptual understanding might be due to a cluster of teaching features rather than just the single highlighted feature. We believe these issues, together with the specific teaching actions that are required and the mechanisms that are responsible for the effects of these teaching features, define promising research agendas for the future. We also believe that they warrant investigation using outcome measures of the re-

---

[4] In brief, mathematical proficiency is the simultaneous and integrated acquisition of five kinds of mathematical competencies, or "*strands:*

- *conceptual understanding*—comprehension of mathematical concepts, operations, and relations
- *procedural fluency*—skill in carrying out procedures flexibly, accurately, efficiently, and appropriately
- *strategic competence*—ability to formulate, represent, and solve mathematical problems
- *adaptive reasoning*—capacity for logical thought, reflection, explanation, and justification
- *productive disposition*—habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy" (National Research Council, 2001, p. 116).

cently defined and more ambitious learning goal of mathematical proficiency (National Research Council, 2001). After commenting about the state of current classroom teaching with respect to these features, we pick up on this claim by identifying a number of principles to guide research agendas that would advance our knowledge of key relationships between teaching and learning.

Absence of Features That Support Conceptual Development from U.S. Mathematics Teaching

Before leaving the section on "Developing Conceptual Understanding," we feel compelled to point out that U.S. mathematics classrooms show a striking absence of the very two features we identified as facilitating conceptual understanding. Ever since U.S. classroom practices have been studied systematically, these two features rarely have been found. Hoetker and Ahlbrand (1969) surveyed early descriptions of classroom teaching, across all subject fields, and reported that, since records have been kept, U.S. school teaching often has focused on lower level skills using a tightly controlled and curtailed question-and-answer routine called *recitation*. A series of studies in the 1970s confirmed this general observation for mathematics classrooms. The authors described classrooms focused on developing routine skills with a basic pattern of teaching that was quite consistent across classrooms (Fey, 1979; National Advisory Committee on Mathematics Education, 1975; Stake & Easley, 1978). Welch (1978) described a typical mathematics lesson this way:

> First, answers were given for the previous day's assignment. A brief explanation, sometimes none at all, was given of the new material, and problems were assigned for the next day. The remainder of the class was devoted to students working independently on the homework while the teacher moved about the room answering questions. The most noticeable thing about math classes was the repetition of this routine (p. 5–6).

Recent surveys have shown that this lesson pattern continues today and that instruction rarely includes explicit attention to conceptual development or engages students in struggling with key mathematical ideas (Hiebert et al., 2003; Rowan et al., 2004; Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999; Weiss, Pasley, Smith, Banilower, & Heck, 2003).

The mathematics Video Study of the 1999 TIMSS provides recent and direct information on the nature of U.S. mathematics teaching (Hiebert et al., 2003). To compare eighth-grade mathematics teaching in the United States with six higher achieving countries in Europe and Asia (Australia, Czech Republic, Hong Kong, Japan, Netherlands, and Switzerland), the 1999 Video Study collected a nationally representative sample of about 100 videotaped lessons from each country. An international team of researchers developed more than 75 codes to analyze the structure and organization of lessons, the mathematical content of the lessons, and the way in which content was worked on during the lessons. Only codes that could be applied with at least 85% interrater agreement were used.

One code from the 1999 Video Study that is especially instructive is the nature of the mathematics problems presented and the way in which the problems were worked on with the students.[5] On the basis of the written materials or the oral presentation by the teacher, all problem statements could be classified into one of three categories defined by the mathematical processes that were apparently intended: using procedures, stating concepts, and making connections. "Making connections" is of most interest here because making connections (among ideas, facts, and procedures) lies at the heart of our definition of conceptual development. A making connections problem could be "Graph the equations $y = 2x + 3$, $2y = x - 2$, and $y = -4x$, and examine the role played by the numbers in determining the position and the slope of the associated lines." Problem statements classified as stating concepts asked students to recall or illustrate definitions or properties. These problems seldom involved conceptual development because they simply asked students to recall or repeat concepts or propositions. A stating concepts problem could be "Show the point (3, 2) on the coordinate plane." Using procedures problems asked students to use apparently known procedures to solve problems, for example, "Solve $2x + 3 = 5$." Results showed that 17% of the problems in an average U.S. lesson indicated an intent to make connections; this was within the range of 15% to 24% evident in all other countries except Japan (Japan was an outlier with 54%).

The picture changes, however, when considering the way in which teachers worked with students on problems. All problems that had some public discussion or visible work were coded a second time into one

[5] This code was developed by Margaret Smith (2000) and built upon earlier work by Mary Kay Stein and colleagues (Stein, Grover, & Henningsen, 1996).

of the same three categories. A fourth category, giving only the answer, was needed for those cases when no mathematical work was done with students except reading the answer. A question of special interest is, "What happened to the making connections problems when they were implemented in the classroom?" In all countries except for Australia and the United States, 37% to 52% of these problems were worked on so that the connections implied by the problem statement were made explicit with the students—through examining the problem, comparing solution methods, justifying why the solution methods worked, and so on. In Australia, 8% of the making connection problems were worked on in this way. In the United States, so few making connection problems were worked on in this way that the percentage rounded to 0% (Hiebert et al., 2003). This nationally representative sample of eighth-grade mathematics lessons shows that students in the United States are not receiving instruction that supports conceptual development.

Before speculating about the reasons for the absence of these teaching features from U.S. classrooms, we point out the degree to which researchers in the Video Study had to probe the actual work of teachers in these classrooms to detect the differences among countries. If coding had stopped with the kinds of mathematics problems presented to students, U.S. classrooms would have looked similar to several other countries. Describing teaching—how students and teachers interact about content—required moving beyond the curriculum to describing the details of how teachers use the curriculum. Documenting connections between teaching and learning requires describing teaching with considerable detail and precision. Notice that this fact calls into question findings that connect opportunity to learn with student achievement when opportunity to learn is defined solely in terms of content coverage (Fletcher, 1971; Gamoran, Porter, Smithson, & White, 1997; Porter et al., 1988). To the extent that teachers transform problems, as they did in the TIMSS 1999 Video Study, the opportunity to learn indicated by an analysis of the curriculum might not be the opportunity to learn experienced by the students. This becomes especially critical if one hopes to understand the mechanism lying behind statistical relationships and develop policy recommendations based on the findings. Simply put, curriculum changes are unlikely to lead directly to changes in students' opportunity to learn.

A number of factors might explain the long-standing absence from U.S. mathematics classrooms of teaching features that support conceptual understanding. These include the difficulty of changing practices that are embedded in a culture, the lack of subject matter and pedagogical knowledge that teachers need in order to teach in different ways, the challenge of enacting complex skills even when they are known by teachers, the limiting nature of common professional development activities, the pressure of district- and state-mandated tests, and the absence of a useful and accessible knowledge base for teachers to improve their practices. We will not develop the arguments that justify each of these explanations, but we will use the last of the explanations, the absence of a knowledge base, as the springboard for the final section of the chapter. The question we tackle in the final section concerns the kind of research agendas that could stimulate further progress in linking teaching and learning. What kind of information will be most helpful and how can it be produced?

FUTURE DIRECTIONS FOR RESEARCH THAT CONNECTS TEACHING AND LEARNING

In this section, we identify several principles that we believe will elevate the quality of research that aims to link mathematics teaching and learning. Some of these principles can be applied immediately, and some will require development themselves. We believe that attending to each of the principles, even if they are not fully developed, will increase the chances that research findings will contribute to a growing, coherent, and useful base of knowledge for improving classroom teaching.

Be Explicit About Learning Goals

The distinction we elaborated earlier, between teaching features that support skill efficiency and those that support conceptual understanding, foreshadows the first principle of conducting research on teaching/learning relationships: Researchers must be clear and explicit about the kinds of learning they will study. Talking about teaching, in general terms, as "effective" or "not effective" is no longer helpful. Teaching is effective (or not) for helping students achieve particular kinds of learning goals. Teaching is effective *for* something and the something must be described and measured as precisely as possible.

We are not saying that a simple correspondence exists between specific features of teaching and specific learning outcomes. This empirical question has yet to be answered. Of course, the data we reviewed earlier indicate already that the correspondence will not be straightforward. The features of teaching that support conceptual understanding might also support skill fluency. Regardless of how these particular relationships are resolved, one lesson from past work is that further progress in understanding connections between teaching and learning depends on being clear about the learning outcomes to be measured and on developing assessments that measure these outcomes directly.

We recommend that the learning goal of mathematical proficiency (National Research Council, 2001) become part of some research agendas in the future. This ambitious goal, which is defined as the integration of five strands of knowledge, has not yet been operationalized through instrument development. Considerable effort will be needed to construct tasks that measure the integration of these strands. Measuring each strand individually misses the point of the construct. It is the integration of the strands that characterizes this promising goal. How to handle these conceptual and methodological challenges presents an immediate task worth pursuing.

Build and Use Theories

Once learning goals are chosen and instruments developed to measure them, theories must guide the research process. Ideally, theories will guide key aspects of the *entire* process: (1) the generation of testable hypotheses about the connections between teaching and learning, (2) the selection and modification of methodological designs that are appropriate for the questions, and (3) the analyses and interpretation of the data. Earlier, we identified the lack of theories of teaching and learning relationships as a major cause for the absence of better claims. Now, we remind the reader that the existence of theories, even local theories, has great benefits for guiding the design of individual studies and for connecting the findings from individual studies to build a useful knowledge base for teaching.

We underscore one benefit of working from theories—they allow researchers to *understand* what they are studying. Theories of teaching provide a framework within which to understand the interactions among features of teaching and the reasons that features of teaching might be more or less effective, within different systems, for facilitating particular kinds of learning. Unless researchers, and teachers, understand why teaching influences learning in a particular way, the knowledge is of little use (Rowan et al., 2002). Understanding the reasons for the effects of teaching allows teachers to adapt the features of interest to their own setting without fatal changes (Brown & Campione, 1996). This makes knowledge useful. When connections between teaching and learning are not understood, isolated findings of "what works" can become rules or prescriptions for practice that are applied too rigidly or changed inappropriately. Users need to know how to modify or adjust the prescriptions to fit the expected differences from classroom to classroom. This benefit of working from theories becomes critical for our discussion of policy in the concluding section of the chapter.

The process through which theories of mathematics teaching and learning relationships are best developed is not yet clear. Ruthven (2002) described two different processes for building theories of teaching, attributed to French and Italian researchers, respectively. A first approach is characterized by top-down activities in which researchers specify the hypotheses to be tested and teachers implement the tests as prescribed. The second approach is characterized by bottom-up activities in which the teachers experiment with various instructional strategies and, together with researchers, detect more general principles that gradually emerge. We believe it is too early in the history of mathematics education research to tell whether one or the other, or a third, approach will be most productive for generating useful theories. All approaches should be considered.

A call to build theories is, fundamentally, a call for researchers to be clear about the hypotheses they use when conducting empirical studies. Hypotheses are researchers' expressions of how well they understand the phenomena under investigation. Explicit statements of hypotheses provide a measure for how the field is advancing in its collective understanding of the phenomena. Thus, hypotheses and the theories they generate not only aid researchers in conducting empirical studies, they provide an essential way for researchers to communicate with one another regarding their own understandings of the teaching-learning relationships. As specific, local hypotheses are made explicit, are refined, and are accumulated, theories can emerge that tie together individual studies and enable the construction of a useful knowledge base.

Set Realistic Expectations for What a Knowledge Base Can Do

A third principle that can inform future research agendas is recognition of the appropriate role of research in setting policy and recommending practice. To define an appropriate role for research on relationships between teaching and learning, consider first Gage's (1978) observations. Gage argued that to develop a science of teaching defined by high predictability and control would be unrealistic. The reasons are found in the challenges presented earlier in documenting straightforward connections between teaching and learning. But, said Gage, a scientific basis for the art of teaching is realistic. It is realistic to expect an increasingly sound theoretical and empirical basis for making informed (although still uncertain) decisions. Gage quoted an 1891 observation by the philosopher Josiah Royce:

> It is vain that the inadequacy of science is made a sufficient excuse for knowing nothing of it. The more inadequate science is when alone, the more need of using it as a beginning when we set about our task" (p. 20).

A realistic expectation, in our view, is that a knowledge base can inform policy and practice and can be used to make decisions with some level of confidence. The level of confidence rises as the knowledge base becomes more detailed, richer, and more coherent. Because research is further advanced in some areas than others, decisions about teaching practices that support some learning goals will be made with more confidence than decisions about achieving other learning goals.

The uncertainty of research claims is quite similar to that found in many other professional fields that involve the complexities of human behavior. Consider the field of health and nutrition. Scientific research on the effects of exercise, diet, and drugs informs recommendations regarding healthy lifestyles. Fruits and vegetables are good for you in a variety of ways. A small daily dose of aspirin reduces the risk of stroke. A high-fiber diet reduces the risk of colon cancer. But, like in education, most of these recommendations are made with some level of confidence, not with certainty. And as the specificity of the recommendation increases, the confidence level declines. Fruits and vegetables are good, but which ones are best, and exactly how many servings should everyone have? Should everyone take a small dose of aspirin each day? What about young people, and what about those who have mild stomach reactions to aspirin? Like education, the problem is that so many interactions influence the effects of any particular treatment—physiology of the individual, amount of exercise, weight, age, and so on. Specifying optimal diet, drug treatments, and lifestyle with certainty is impossible. But a knowledge base that informs decisions with increasing levels of confidence can be built.

One final parallel between decisions about physical health and about education is important. Readers surely have noticed that, as new information about nutrition and drug treatments becomes available, recommendations change, sometimes dramatically. Should you eat butter or margarine? For years, margarine was the healthy alternative to butter. With new knowledge about trans-fatty acids, butter might be the better choice. Recommendations regarding more serious issues, like hormone treatments and breast cancer, also have witnessed considerable changes, even reversals. This does not mean that the medical profession is being silly or lazy or is conducting thoughtless and flawed research. Rather, the problems are extremely complicated, and sorting out all of the interactions and conditions under which treatments affect outcomes is enormously difficult. It also is noteworthy that people generally believe they benefit from the latest and currently best information that the health and medical professions can provide, even if it will change in the future. The same should be true in education with respect to relationships between teaching and learning.

One implication of this analysis is that educators, and the public, should not expect critical experiments in education. They should not expect single studies, regardless of how large or how well conceived, to answer the question of what kind of teaching is best for achieving particular learning goals. Well-designed studies that yield clear results can increase the level of confidence about which features of teaching to recommend, but the recommendations always are subject to change when better information becomes available or when conditions change.

We conclude this section by reminding the reader that not all decisions in education are, or should be, based on research. Learning goals, for example, are statements about what society most values and should be selected, in part, through public debates about values (Hiebert, 2003). How important is conceptual understanding? Should it be a priority or are other competencies more important? Research can inform society about what is possible but not about what is most valued. Once learning goals have been selected and made explicit, based at least partly on value judgments, then researchers can examine the features of teaching that facilitate the achievement of such goals.

Account for the Costs and Benefits of Different Research Approaches

The fourth principle we propose to elevate the quality of research on classroom teaching and learning relationships is to carefully consider the advantages and disadvantages of different research designs and strategies when conducting empirical studies and interpreting data. We elaborate this principle by considering the issues related to three particular contexts: research designs that compare the effects of instructional methods; research designs that correlate features of teaching with student learning; and the trade-offs between small-scale qualitative studies and large-scale quantitative studies.

Comparing the Effects of Different Instructional Methods

Research designs that compare students' learning under two or more instructional methods often are quasi-experiments (Campbell & Stanley, 1963). Some of the methodological features are consistent with usual conventions of scientific experiments (controlling some variables while studying differences among others) but some are not (lack of random assignment of students to instructional method). One advantage of these comparative studies is that they often have some ecological validity. That is, they are conducted in real classrooms with all of the usual complexities. Instructional methods are implemented as systems, with all of the usual interactions. A second advantage is that these studies have the opportunity to control some of the most obvious alternative explanations for the results. For example, pretests can show similar entry knowledge and skills by students across treatment groups, content can be similar across groups except in ways dictated by the different teaching methods, and classroom observations can ensure faithful implementation of methods.

Some of the disadvantages of the quasi-experiments that compare teaching methods are inherent in the design and are difficult for researchers to control. The ecological validity that assures a realistic implementation of a teaching method also guarantees that the method will not be exactly replicable in other settings. As noted earlier, many classroom conditions, including the students, influence the way in which methods are implemented and, consequently, influence the learning outcomes. The challenge for researchers is to determine whether robust features of teaching within the system yield particular kinds of outcomes regardless of the naturally occurring changes in the system from classroom to classroom. This requires multiple replications, a strategy not frequently employed by educational researchers. But replication is one of the most powerful tools available to sort out which features of teaching make a difference across different implementations of a system of teaching, and even across different systems.

Another disadvantage of these comparative studies of teaching is easier for researchers to address. Reports of comparative studies often provide insufficient descriptions of the teaching features or methods studied. Without full descriptions of the teaching features being compared, the results cannot be interpreted in ways that connect particular kinds of teaching with particular kinds of learning. And without these kinds of interpretations, it is impossible to *understand* the relationships that are documented. Superior learning reported during a thinly described "student-centered" approach compared with a thinly described "direct teaching" approach is of no value. There is no way to develop more specific hypotheses about what, exactly, in the student-centered approach facilitated students' learning. And it is these ill-understood connections that can lead to "lethal mutations" when implementing the teaching features in other conditions (Brown & Campione, 1996).

Thinly described teaching methods often are special deficiencies in reports of studies that include a control condition. The experimental method usually is described in some detail, but the control method can be described as cryptically as "traditional instruction was provided" or "the textbook was followed closely." Thinly described control methods undermine the educational significance of a claim that the experimental method yielded better results. This is because the power of such a claim depends as much on the quality of the control method as the quality of the experimental method. When benefits are claimed for one teaching method over another, regardless of the methods being compared, the claim is only as educationally significant as the quality of the weaker method. If the weaker method is extremely weak, or not sufficiently described to judge its quality, then results showing that the alternative method yielded better results are of little importance. To evaluate comparative claims between teaching methods, all the methods must be fully described.

We conclude this section with a suggestion about research questions. History shows that hypotheses in social science research are almost always confirmed (Greenwald, Pratkanis, Lieppe, & Baumgardner, 1986). This is especially true for predictions regarding the superiority of an experimental condition over a control condition. Several well-known research biases, such as the Hawthorne effect, probably explain the frequent confirmation of hypotheses. But, as Greenwald and colleagues (1986) argued, this tendency can and should be countered by changing the nature of the research question. Rather than asking "Is Method A better than Method B?" researchers should ask "Under what conditions is Method A better than Method B?" Translating the suggestion into the situation of interest here, we recommend asking more questions of the kind "Under what conditions will teaching features of Set A facilitate achievement of particular mathematics learning goals more effectively than teaching features of Set B?" Indeed, this is exactly the characteristic of Brownell and Moser's (1949) study that received applause from Cronbach (1986). Cronbach echoed Greenwald and colleagues' (1986) recommendation by noting the benefits of documenting the conditions under which particular instruction approaches are effective rather than claiming one approach is superior to another.

Correlating Features of Teaching with Students' Learning

A common research approach to connecting teaching with learning has been to measure a range of specific features of teaching and gains in students' achievement and then, through correlation techniques, to identify the features of teaching that correspond best with students' learning. Known as *process-product research* (Brophy & Good, 1986; Dunkin & Biddle, 1974), this approach often focuses on counting the frequency of teacher behaviors and correlating these with student gains on standardized achievement tests. The approach was primarily responsible for the patterns described earlier between features of teaching and students' skill efficiency.

Advantages of the correlation approach center around the ability to sort out features of teaching that might be key facilitators of learning and those that are part of the implementation but might not be critical. For example, teaching might vary along dimensions of asking students to present alternative solution methods and time spent working in groups developing shared solution methods. One of these features (or an interaction between the two) might be more important than the other for students' conceptual development. Well-designed correlation research could provide useful information. Schoen, Cebulla, Finn, and Fi (2003) presented an example of how correlations can be used to sort out key instructional variables in classrooms using reform curricula. Although the traditional process-product approach selected a specific kind of teacher behavior and used standardized tests to measure learning, the correlational approach is not limited to these independent and dependent variables. The approach is especially useful to create initial maps of the terrain when few hypotheses exist to suggest which features of teaching might facilitate which kinds of learning.

A potential danger of the correlation approach is the flip-side of its greatest advantage—the assumption that teaching consists of a collection of interchangeable features and that the effects of individual features can be measured independently of their interactions with other features in the system. Although many researchers have been explicit that "any teaching act is meaningful only in relationship to other instructional behaviors that occur in the classroom" (Good et al., 1983, p. 20), their warnings have often been overlooked in summaries and syntheses of research and when research findings are incorporated into professional development activities for teachers. Without considering the systems of which features are a part, recommendations from process-product relationships can be as simple and misguided as "Teachers should engage in process [X] more often." Thus, as we noted earlier, when discussing the challenges faced in developing adequate theories of teaching and learning relationships, teaching must be conceived as a system of interacting features and overly simplistic notions of teaching must be avoided.

In spite of the danger, we believe that correlation approaches have a role to play in linking teaching with learning. They can help to identify features of teaching that might have major effects on learning and deserve further study, and they can provide descriptive data to inform the development of hypotheses regarding teaching-learning relationships. Equally important, especially in the complex classroom environment where school learning takes place, correlational studies can assist with decisions on which variables will likely play major roles and which will play minor roles in particular learning situations and in this way help develop testable hypotheses and assist with the design of manageable experiments.

Balancing the Benefits of Small-Scale Qualitative Studies and Large-Scale Quantitative Studies

In the current research climate, there are increasing pressures to quantify regularities through large-scale experimental studies that use random assignment and seek to sort out cause and effect between classroom practice and student learning (Coalition for Evidence-Based Policy, 2003). Partly in response to these pressures, the National Research Council sponsored a report that proposed guidelines for what counts as "scientific" research in education (National Research Council, 2002). A number of researchers have argued that a balanced approach between qualitative and quantitative approaches has the best chance of moving the field forward (see, for example, Maxwell, 2004). We believe that small-scale qualitative approaches and large-scale quantitative approaches and mixes between the two all have useful roles to play in advancing our understanding of the relationships between teaching and learning. Gage's (1978) simple observation is a good starting point for outlining the contributions of different approaches: qualitative approaches reveal what is possible and quantitative approaches demonstrate what is probable. But we can be more specific.

Qualitative and small-scale studies can provide a number of critical ingredients for the research enterprise. Such studies are essential for developing and testing measuring instruments for teaching and learning, including the high-quality measures of student understanding of complex mathematical ideas called for in the National Research Council (2004) report on the effectiveness of the evaluations of mathematics reform curricula. Qualitative and small-scale studies also can support the thorough development of instructional treatments before large-scale studies test their effects with larger samples of students. Additionally, although case studies are not frequently undertaken to test hypotheses, they can, as Yin (2000) argued, rule out important alternate hypotheses if the studies are carefully designed.

Finally, and perhaps most importantly, qualitative smaller scale studies can provide data that deepen our understanding of how systems of teaching work to facilitate learning (Maxwell, 2004). Another brief story from the medical field is instructive. As reported during the December 7, 2004 National Public Radio program "Morning Edition" (Knox, 2004), a very small-scale study had recently tested the effect of a vaccine targeting a pre-leukemia condition called myelodysplasia. The vaccine not only induced immune responses in about 60% of the cases but unexpectedly created remissions in several cancer cases. "Now the challenge," according to Richard Knox, "is to figure out why the vaccine works." Until the researchers understand the reasons for the vaccine's effect, they cannot adapt it to trials under other conditions or on large scales. There are good reasons to apply a similar logic to research on teaching and learning. Qualitative and small-scale studies can detect the existence of teaching effects and then unpack the system of interacting teaching features to explore the reasons for the effects. Understanding why particular features of teaching facilitate particular kinds of learning is essential for creating more refined hypotheses, for adapting the teaching system to other conditions, and for testing its effects on a large scale.

Large-scale quantitative studies play an equally critical role in building a useful knowledge base for teaching. Because the links between teaching and learning are mediated by so many contextual variables, large-scale studies are needed to gauge the size of these effects and estimate the likelihood of finding similar links between teaching and learning across multiple classrooms. Researchers cannot control all of the classroom influences on students' learning and all of the interaction effects among the teaching features of interest and the systems in which they operate, so replication over large numbers of classrooms and students remains one of the best strategies for detecting links between teaching and learning (Hiebert, Gallimore, & Stigler, 2002). We recommend taking seriously the call for large-scale cause-effect studies that aim to test hypotheses about relationships between teaching and learning. Along with Cohen et al. (2003) and others before them (e.g., Brownell & Moser, 1949), we recommend studying the effects of implementing systems of teaching under different conditions with clear learning goals in mind. To the extent that conditions not of interest can be controlled (e.g., instruction implemented similarly across classrooms, random assignment of students to treatments), the results can be interpreted with clarity and confidence. As we argued earlier, full control is unlikely but need not preclude efforts to design studies with appropriate controls.

Smaller scale qualitative studies and larger scale quantitative studies are essential for building a theoretically driven and practically useful knowledge base for teaching. Neither is sufficient; both are necessary. The confidence with which conclusions can be drawn about the effectiveness of features or systems of teaching increases with the convergence toward similar conclusions from studies using different methods. We agree with the National Research Council (2002), "If a research conjecture or hypothesis can withstand the scrutiny of multiple methods its credibility is enhanced greatly" (p. 64). The issue is not one of choosing between approaches but rather one of using both approaches in a mutually beneficial way.

FINAL THOUGHTS

We began the chapter with the goal of examining directly the issue of teaching effectiveness—why it has been so hard to document, what is known about it, and how the mathematics education community can learn more about the relationships between teaching and learning. We contended that these questions get to the heart of the educational research enterprise because (1) the questions lie at the foundation of education activity and are fundamental research issues and (2) the improvement of classroom practice depends on the answers. This places researchers in a difficult position. The issues they are dealing with are fundamental and complex, but researchers are under pressure from policy and government agencies to produce simple, easy-to-understand evidence-based answers and to produce them quickly. What constitutes an appropriate response to this challenging dilemma?

One response to the challenge, currently popular in the United States, is to address the simple question "What works?" The "What Works Clearinghouse" launched in 2004 (Viadero, 2004) is the latest in a series of efforts by the U.S. government (U.S. Department of Education, 1987), state governments (e.g., Dixon et al., 1998), and professional organizations (e.g., Cawelti, 2004) to provide practitioners with straightforward evidence-based advice for improving classroom practice. The What Works Clearinghouse plans to roll out a series of recommendations on a range of educational issues by conducting reviews of intervention programs, practices, and products. Similar to the review procedure used by Dixon et al. (1998) to identify best practices in mathematics classrooms, the What Works Clearinghouse will screen out most studies for review using a set of selection criteria. In particular, all studies that do not include a cause-effect analysis through the use of comparison groups will be excluded. The studies remaining constitute one segment of the large-scale quantitative studies described earlier. These studies are an important part of the picture. They provide essential information regarding teaching-learning relationships. But, usually, these large-scale comparative studies do not help consumers understand why the documented relationships exist or exactly how particular kinds of teaching facilitated particular kinds of learning. As we have argued, users must understand why particular features of teaching facilitate students' learning in order to deal appropriately with the modifications that inevitably are needed for their own classrooms. Some level of understanding is essential to prevent lethal mutations in teaching (Brown & Campione, 1996).

We applaud the concept of conducting and reviewing research to answer questions of what works. Educators need trustworthy information to inform classroom decisions. Our concerns about the limitations of the process engaged by the What Works Clearinghouse in no way lessen the importance or urgency of the goal toward which the Clearinghouse is working. In fact, we believe this goal must be aggressively pursued by researchers in the mathematics education community. Mathematics education researchers are in the best position to create and sustain a better alternative.

We conclude this chapter by proposing a richer, more useful, but longer term alternative to "what works." We propose the coherent and systematic construction of a knowledge base that documents robust links between teaching and learning *and* provides insights into the mechanisms that are responsible for such links. Researchers will need to participate in efforts to generate knowledge about teaching-learning relationships that have direct relevance for classroom teachers and to regularly review available evidence to provide recommendations accompanied with estimates of confidence. We expect that these efforts will follow a range of different models for researcher-teacher collaboration (Hiebert et al., 2002; Ruthven, 2002) but will be characterized by the principles outlined in this chapter. Effective teaching will be studied in relation to clearly specified and measured learning outcomes so that conclusions will be phrased, not just as "effective teaching" but as "effective teaching under conditions [X] for learning [Y]." Hypotheses and theories will guide all aspects of the research process so that conjectures can be offered about why and how the targeted teaching features or systems facilitated (or did not facilitate) students' learning. The conditions under which features were effective will be detailed so that users can interpret the conclusions in light of their own local situations. Tests of hypotheses about teaching-learning relationships will be replicated in a variety of contexts to study the influence of contextual factors so that conclusions can be drawn with increasing specificity and confidence. Reviews will look for patterns of consistent findings from studies using different research methods to increase the confidence with which conclusions can be offered.

To build the richer knowledge base we envision, debates about research methods must shift from debates about which methods are appropriate to debates about how to increase the quality of the methods employed. The principle that multiple methods are better than one method must be accepted, and researchers will need to pursue the goal of gathering data with higher levels of trustworthiness, regardless of method. Discussions like those initiated by Hanna (1998), Kilpatrick (1993), Lester and Lambdin (1998), Nissen and Blomhøj (1993), and Simon (2004) on the quality of research in mathematics education should continue and be extended.

A knowledge base of the kind we envision cannot be built without resources. Conducting the large-scale comparisons of teaching that provide critical kinds of information are especially expensive. Assembling networks of researchers who can launch coordinated series of studies and build on one another's work requires long-term investments. No research enterprise that produces important advances is cheaply funded. U.S. mathematics education research funding comes primarily from the National Science Foundation and the U.S. Department of Education, and, as a Rand-conducted study of funding for mathematics education research from 1996 to 2001 pointed out, neither agency has clear strategic long-range planning for mathematics education R&D funding. In fact, funding for research and development in both agencies is a miniscule part of their total education budgets (Lacampagne, Newton, & Markham, 2001). Apparently high-quality evidence-based information on teaching-learning relationships in the mathematics classroom will cost much more than the United States currently is willing to spend.

The knowledge base we envision that links mathematics teaching and learning is beginning to take shape, even with limited resources. But it is difficult to find. It is hidden in the archives of libraries and scattered across time and among countless materials. Teachers (and researchers) must work hard to access it. Recent reviews have succeeded in pulling together some information (English, 2002; Kilpatrick et al., 2003; National Research Council, 2001; this volume), and we hope this chapter extends and contributes to these efforts. But most of this work is post hoc, requiring reviewers to detect patterns across work that was conducted independently and without connections. Much can be gained by researchers launching research agendas that are intentionally and explicitly connected. Through theory development, planned replications, and other research activities that connect research communities, a coherent, rich, accessible, and useful knowledge base can be built. Questions of "What works?" can then be answered with the richness and insight teachers deserve.

REFERENCES

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning.* New York: Grune & Stratton.

Ball, D. L. (1993). Halves, pieces, and twoths: Constructing and using representational contexts in teaching fractions. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 157–195). Hillsdale, NJ: Erlbaum.

Ball, D. L., & Rowan, B. (2004). Introduction: Measuring instruction. *Elementary School Journal, 105,* 3–10.

Berliner, D. C. (1976). Impediments to the study of teacher effectiveness. *Journal of Teacher Education, 27,* 5–13.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education, 29,* 41–62.

Borko, H., Stecher, B. M., Alonzo, A., Moncure, S., & McClam, S. (2003). *Artifact packages for measuring instructional practice: A pilot study.* Retrieved June 30, 2004, from http://www.cresst.org/reports/R615.pdf

Brophy, J. E. (1988). Research on teacher effects: Uses and abuses. *Elementary School Journal, 89,* 3–21.

Brophy, J. E. (1997). Effective instruction. In H. J. Walberg & G. D. Haertel (Eds.), *Psychology and educational practice* (pp. 212–232). Berkeley, CA: McCutchan.

Brophy, J. E. (1999). *Teaching* (Education Practices Series No. 1). Geneva: International Bureau of Education. Retrieved June 10, 2002, from http://www.ibe.unesco.org

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.

Brousseau, G. (1997). *Theory of didactical situations in mathematics.* Dordrecht, The Netherlands: Kluwer.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Erlbaum.

Brown, S. I. (1993). Towards a pedagogy of confusion. In A. M. White (Ed.), *Essays in humanistic mathematics* (pp. 107–121). Washington, DC: Mathematical Association of America.

Brownell, W. A. (1935). Psychological considerations in the learning and teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic: Tenth yearbook of the National Council of Teachers of Mathematics* (pp. 1–31). New York: Teachers College, Columbia University.

Brownell, W. A. (1947). The place of meaning in the teaching of arithmetic. *Elementary School Journal, 47,* 256–265.

Brownell, W. A. (1948). Criteria of learning in educational research. *Journal of Educational Psychology, 39,* 170–182.

Brownell, W. A., & Moser, H. E. (1949). Meaningful vs. mechanical learning: A study in grade III subtraction. *Duke University Research Studies in Education* (No. 8). Durham, NC: Duke University Press.

Brownell, W. A., & Sims, V. M. (1946). The nature of understanding. In N. B. Henry (Ed.), *Forty-fifth yearbook of the National Society for the Study of Education: Part I. The measurement of understanding* (pp. 27–43). Chicago: University of Chicago Press.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Washington, DC: American Educational Research Association.

Capon, N., & Kuhn, D. (2004). What's so good about problem-based learning? *Cognition and Instruction, 22,* 61–79.

Carpenter, T. P., Fennema, E., Peterson, P. L, Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26,* 499–531.

Carroll, J. (1963). A model for school learning. *Teachers College Record, 64,* 723–733.

Cawelti, G. (Ed.). (2004). *Handbook of research on improving student achievement.* (3rd ed.). Arlington, VA: Educational Research Service.

Coalition for Evidence-Based Policy. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved August 5, 2004, from http://www.ed.gov/rschstat/research/pubs/rigorousevid/index.html

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematics development. *Educational Researcher, 23*(7), 13–20.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problemcentered second-grade mathematics project. *Journal for Research in Mathematics Education, 22,* 3–29.

Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (Eds.). (1993). *Teaching for understanding: Challenges for policy and practice.* San Francisco: Jossey-Bass.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25,* 119–142.

Cooney, T. J. (1985). A beginning teacher's view of problem solving. *Journal for Research in Mathematics Education, 16,* 324–336.

Cronbach, L. J. (1986). Social inquiry by and for earthlings. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science: Pluralisms and subjectivities* (pp. 83–107). Chicago: University of Chicago Press.

Davis, R. B. (1984). *Learning mathematics: The cognitive science approach to mathematics education.* Norwood, NJ: Ablex.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Dewey, J. (1910). *How we think.* Boston: Heath.

Dewey, J. (1926). *Democracy and education.* New York: Macmillan.

Dewey, J. (1929). *The quest for certainty.* New York: Minton, Balch & Co.

Dixon, R. C., Carnine, D. W., Lee, D.-S., Wallin, J., The National Center to Improve the Tools of Educators, & Chard, D. (1998). *Review of high quality experimental mathematics research: Report to the California State Board of Education.* Eugene, OR: National Center to Improve the Tools of Educators.

Doyle, W. (1978). Paradigms for research on teacher effectiveness. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 163–198). Itasca, IL: Peacock.

Duffy, G. G. (1981). Teacher effectiveness research: Implications for the reading profession. In M. L. Kamil (Ed.), *Directions in reading: Research and instruction. Thirtieth yearbook of the National Reading Conference* (pp. 113–136). Washington, DC: National Reading Conference.

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

English, L. (Ed.). (2002). *Handbook of international research in mathematics education.* Mahwah, NJ: Erlbaum.

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American Educational Research Journal, 17*(1), 43–60.

Fawcett, H. P. (1938). *The nature of proof: A description and evaluation of certain procedures used in a senior high school to develop an understanding of the nature of proof.* New York: Teachers College, Columbia University.

Fennema, E., & Romberg, T. A. (Eds.). (1999). *Mathematics classrooms that promote understanding.* Mahwah, NJ: Erlbaum.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Fey, J. (1979). Mathematics teaching today: Perspectives from three national surveys. *Mathematics Teacher, 72,* 490–504.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahn, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 7–32). Washington, DC: U.S. Department of Health, Education, and Welfare, National Institute of Education.

Fletcher, H. J. (1971). An efficiency reanalysis of the results. *Journal for Research in Mathematics Education, 2,* 143–156.

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* (4th ed., pp. 3–16). New York: Macmillan.

Floden, R. E. (2002). The measurement of opportunity to learn. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 231–266). Washington, DC: National Academy Press.

Forman, E. A. (2003). A sociocultural approach to mathematics reform: Speaking, inscribing, and doing mathematics within communities of practice. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 333–352). Reston, VA: National Council of Teachers of Mathematics.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht, The Netherlands: Reidel.

Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education, 21,* 180–206.

Gage, N. L. (1978). *The scientific basis of the art of teaching.* New York: Teachers College Press.

Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart and Winston.

Gamoran, A. (2001). Beyond curriculum wars: Content and understanding in mathematics. In T. Loveless (Ed.), *The great curriculum debate: How should we teach reading and math?* (pp. 134–162). Washington, DC: Brookings Institution Press.

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,* 325–328.

Good, T. L., & Grouws, D. A. (1977). A process-product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28*(3), 49–54.

Good, T. L., & Grouws, D. A. (1979). The Missouri mathematics effectiveness project: An experimental study in fourth-grade classrooms. *Journal of Educational Psychology, 71* (3), 355–362.

Good, T. L., Grouws, D. A., & Ebmeier, H. (1983). *Active mathematics teaching.* New York: Longman.

Good, T. L., Mulryan, C., & McCaslin, M. (1992). Grouping for instruction in mathematics: A call for programmatic research on small-group processes. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 165–196). New York: Macmillan.

Gravemeijer, K. (1994). *Developing realistic mathematics education.* Culemborg, The Netherlands: Technipress.

Greenwald, A., Pratkanis, A., Lieppe, M., & Baumgardner, M. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93,* 216–229.

Grouws, D. A. (2004). Mathematics. In G. Cawelti (Ed.), *Handbook of research on improving student achievement* (3rd ed., pp. 160–178). Arlington, VA: Educational Research Service.

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25,* 1–29.

Handa, Y. (2003). A phenomenological exploration of mathematical engagement: Approaching an old metaphor anew. *For the Learning of Mathematics, 23,* 22–28.

Hanna, G. (1998). Evaluating research papers in mathematics education. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics as a research domain: A search for identity* (ICMI Studies Series, Vol. 4, pp. 399–407). Dordrecht, The Netherlands: Kluwer.

Hatano, G. (1988). Social and motivational bases for mathematical understanding. In G. B. Saxe & M. Gearhart (Eds.), *Children's mathematics* (pp. 55–70). San Francisco: Jossey-Bass.

Heath, R. W., & Nielson, M. A. (1974). The research basis for performance-based teacher education. *Review of Educational Research, 44,* 463–484.

Heaton, R. M. (2000). *Teaching mathematics to the new standards: Relearning the dance.* New York: Teachers College Press.

Heid, M. K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education, 19,* 3–25.

Hiebert, J. (2003). What research says about the NCTM Standards. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 5–23). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25*(4), 12–21.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M-Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. W. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study* (NCES 2003-013). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*(5), 3–15.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *Elementary School Journal, 101,* 3–20.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30,* 393–425.

Hoetker, J., & Ahlbrand, W. P., Jr. (1969). The persistence of the recitation. *American Educational Research Journal, 6,* 145–167.

Husen, T. (1967). *International study of achievement in mathematics: A comparison of twelve countries (Vols. 1–2).* New York: John Wiley.

Inagaki, K., Hatano, G., & Morita, E. (1998). Construction of mathematical knowledge through whole-class discussion. *Learning and Instruction, 8,* 503–526.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition, 3,* 192–212.

Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eighth-grade mathematics classrooms in Germany, Japan, and the United States. *Mathematical Thinking and Learning, 1*(4), 255–278.

Kilpatrick, J. (1993). Beyond face value: Assessing research in mathematics education. In G. Nissen & M. Blomhøj (Eds.), *Criteria for scientific quality and relevance in the didactics of mathematics* (pp. 15–34). Roskilde, Denmark: Danish Research Council for the Humanities.

Kilpatrick, J., Martin, W. G., & Schifter, D. (Eds.). (2003). *A research companion to Principles and Standards for School Mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Knox, R. (Speaker). (2004, December 7). *New report suggests that researchers have had some success in making a cancer vaccine* [Radio broadcast]. Washington, DC: National Public Radio.

Lacampagne, C.B., Newton, E., & Markham, K. (2001). Federal funding of mathematics education research. Project Memorandum (PM-1238) Prepared for Office of Educational Research and Improvement (OERI). Arlington, VA: RAND.

Lampert, M. (2001). *Teaching problems and the problems of teaching.* New Haven, CT: Yale University Press.

Lampert, M., & Cobb, P. (2003). Communication and language. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 237–249). Reston, VA: National Council of Teachers of Mathematics.

Lave, J. (1988). *Cognition in practice.* Cambridge, U.K.: Cambridge University Press.

Leinhardt, G. (1986). Expertise in math teaching. *Educational Leadership, 43(7),* 28–33.

Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*(2), 75–95.

Lester, F. K., Jr., & Lambdin, D. (1998). The ship of Theseus and other metaphors for thinking about what we value in mathematics education research. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics as a research domain: A search for identity* (ICMI Studies Series, Vol. 4, pp. 415–425). Dordrecht, The Netherlands: Kluwer.

Malara, N. A., & Zan, R. (2002). The problematic relationship between theory and practice. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 553–580). Mahwah, NJ: Erlbaum.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3–11.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21,* 29–45.

McDonald, F. (1976). Report on phase II of the beginning teacher evaluation study. *Journal of Teacher Education, 27,* 39–42.

McDonald, F., & Elias, P. (1976). The effects of teaching performance on pupil learning. Beginning Teacher Evaluation Study, Phase II, 1974–1976 (Final report, 5 vols.). Princeton, NJ: Educational Testing Service.

National Advisory Committee on Mathematics Education. (1975). *Overview and analysis of school mathematics, grades K–12.* Washington, DC: Conference Board of the Mathematical Sciences.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. Washington, DC: Department of Education.

National Research Council. (1999). *How people learn: Brain, mind, experience, and school*. J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.). Committee on Developments in the Science of Learning, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2002). *Scientific research in education*. R. J. Shavelson & L. Towne (Eds.). Committee on Scientific Principles for Educational Research, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2004). *On evaluating curricular effects: Judging the quality of K–12 mathematics evaluations*. J. Confrey & V. Stohl (Eds.). Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curricular Materials. Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nissen, G., & Blomhøj, M. (Eds.). (1993). *Criteria for scientific quality and relevance in the didactics of mathematics*. Roskilde, Denmark: Danish Research Council for the Humanities.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects. *Educational Evaluation and Policy Analysis, 26,* 237-257.

Olson, L. (1997, January 22). Keeping tabs on quality. *Education Week Supplement*, pp. 7–11, 14–17.

Oser, F. K., & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1031–1065). New York: Macmillan.

Paige, R. (2004). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality, 2004.* Retrieved October 5, 2004, from http://www.title2.org/secReport04.htm

Piaget, J. (1960). *The psychology of intelligence*. Totowa, NJ: Littlefield, Adams.

Polya, G. (1957). *How to solve it* (2nd ed.). Garden City, NY: Doubleday Anchor Books.

Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. A. Grouws, T. J. Cooney, & D. Jones (Eds.), *Effective mathematics teaching* (pp. 96–113). Reston, VA: National Council of Teachers of Mathematics.

Powell, M. (1980). The beginning teacher evaluation study: A brief history of a major research project. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 1–5). Washington, DC: National Institute of Education.

Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.

Reusser, K. (2001, September). *Bridging instruction to learning: Where we come from and where we need to go*. Paper presented at the Ninth Conference of the European Association for Research on Learning and Instruction, Fribourg, Switzerland.

Romberg, T. A., & Carpenter, T. P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 850–873). New York: Macmillan.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmillan.

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education, 34,* 344–363.

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* (CPRE Research Report Series PR-051). Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Rowan, B., Harrison, D. & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *Elementary School Journal, 105,* 103–127.

Ruthven, K. (2002). Linking researching with teaching: Towards a synergy of scholarly and craft knowledge. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 581–598). Mahwah, NJ: Erlbaum.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: The University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system: A quantitative outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousands Oaks, CA: Corwin Press.

Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education, 34,* 228-259.

Schoenfeld, A. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

Schoenfeld, A. (1988). When good teaching leads to bad results: The disasters of "well taught" mathematics courses. *Educational Psychologist, 23,* 145–166.

Schoenfeld, A. (1998). Toward a theory of teaching-in-context. *Issues in Education: Contributions from Educational Psychology, 4,* 1–94.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4–13.

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.

Silver, E. A., & Stein, M. K. (1996). The QUASAR Project: The "revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education, 30,* 476–522.

Simon, M. A. (2004). Raising issues of quality in mathematics education research. *Journal for Research in Mathematics Education, 35,* 157–163.

Simon, M. A., Tzur, R., Heinz, K., & Kinzel, M. (2004). Explicating a mechanism for conceptual learning: Elaborating the construct of reflective abstraction. *Journal for Research in Mathematics Education, 35,* 305–329.

Skemp, R. R. (1971). *The psychology of learning mathematics.* Middlesex, England: Penguin.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Arithmetic Teacher, 26*(3), 9–15.

Smith, M. (2000). *A comparison of the types of mathematics tasks and how they were completed during eighth-grade mathematics instruction in Germany, Japan, and the United States.* Unpublished doctoral dissertation, University of Delaware, Newark.

Stake, R., & Easley, J. (Eds.). (1978). *Case studies in science education.* Urbana: University of Illinois.

Stecher, B., & Borko, H. (2002). Integrating findings from surveys and case studies: Examples from a study of standards-based educational reform. *Journal of Education Policy, 17,* 547–569.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33,* 455–488.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50–80.

Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist, 35,* 87–100.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States.* NCES 1999-074. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom.* New York: Free Press.

Thorndike, E. L. (1906). *The principles of teaching based on psychology.* New York: A. G. Seiler.

Thorndike, E. L. (1912). *Education: A first book.* New York: Macmillan.

Traub, J. (2002, November 10). Does it work? *New York Times.* Retrieved November 18, 2002, from http: //www.nytimes.com/2002/11/10/edlife/10CHILD.html

U.S. Department of Education. (1987). *What works: Research about teaching and learning* (2nd ed.). Washington, DC: U.S. Government Printing Office.

Viadero, D. (2004, July 14). "What Works" research site unveiled. *Education Week, 23*(42), 1, 33.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.). Cambridge, MA: Harvard University Press.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22,* 366–389.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87,* 406–423.

Weinert, F. E., Schrader, F.-W., & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research, 13*, 895–914.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge, U.K.: Cambridge University Press.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States.* Retrieved July 19, 2004, from http://horizon-research.com

Welch, W. (1978). Science education in Urbanville: A case study. In R. Stake & J. Easley (Eds.), *Case studies in science education* (pp. 5-1–5-33). Urbana: University of Illinois.

Winne, P. H., & Marx, R. W. (1980). Matching students' cognitive responses to teaching skills. *Journal of Educational Psychology, 72,* 257–264.

Wittrock, M. C. (1986). Students' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 297–314). New York: Macmillan.

Wubbels, T., Brekelmans, M., & Hooymayers, H. (1992). Do teacher ideals distort the self-reports of their instructional behavior? *Teaching and Teacher Education, 8,* 47–58.

Yin, R. K. (2000). Rival explanations as an alternative to reforms as "experiments." In L. Bickman (Ed.), *Validity & social experimentation: Donald Campbell's legacy* (pp. 239–266). Thousand Oaks, CA: Sage.

AUTHOR NOTE

We thank a number of individuals for their helpful comments on earlier drafts of the chapter. Hilda Borko and Thomas Cooney provided insightful and constructive comments at two critical points during the development of the chapter. Faculty and doctoral students in mathematics education at the University of Delaware offered two rounds of comments during lively seminar discussions. Colleagues at Michigan State University shared generously of their time and advice during a colloquium presentation of these ideas. Finally, we thank Frank Lester, Jr., for providing helpful feedback and coordinating the review of the chapter.