The Link Between Teacher Quality and Student Outcomes: A Research Synthesis

October 2007

Laura Goe, Ph.D.
ETS

1100 17th Street N.W., Suite 500
Washington, DC 20036-4632
877-322-8700 • 202-223-6690
www.ncctq.org

Copyright © 2007 National Comprehensive Center for Teacher Quality, sponsored under government cooperative agreement number S283B050051. All rights reserved.

This work was originally produced in whole or in part by the National Comprehensive Center for Teacher Quality with funds from the U.S. Department of Education under cooperative agreement number S283B050051. The content does not necessarily reflect the position or policy of the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by the federal government.

The National Comprehensive Center for Teacher Quality is a collaborative effort of Education Commission of the States, ETS, Learning Point Associates, and Vanderbilt University.

2234_10/07

Acknowledgments

The author wishes to thank Leslie Stickler, research assistant at ETS, for invaluable assistance in locating and summarizing studies for consideration for this synthesis, as well as useful suggestions during the review and revision process.

The author also would like to thank the following for their helpful reviews of the manuscript: Michael Allen, Jane Coggshall, Tricia Coulter, Stephanie Stoll Dalton, Carol Dwyer, Jon Fullerton, Amy Jackson, Bonnie Jones, and Becky Smerdon.

Contents

Executive Summary

Federal law emphasizes the need for states and districts to ensure that *all* students—particularly at-risk students, minority students, and students in high-poverty areas—have access to highly qualified, experienced teachers. But is it sufficient for a teacher to have "paper" qualifications and teaching experience? After all, appropriate degrees, certification, and experience may be important at a minimum, but do they guarantee quality teaching? This synthesis of the research on teachers and their contribution to student achievement found that a number of studies cite a few areas of teacher quality in which research shows convincingly what matters, whereas the inconsistency of other findings indicates that much is still to be learned.

Key Questions

The key questions covered in this research synthesis are as follows: What is teacher quality? How can it be measured? How important is it to student learning? Do certain aspects of teacher quality have a stronger impact on student achievement for specific students, subjects, or grade levels? How important is teacher experience? How can teacher quality be better understood?

What Is Teacher Quality?

Teacher quality has been defined and measured in many ways. There is nearly universal agreement that teacher quality matters in terms of student achievement, but there has been no clear consensus on which aspects of teacher quality matter *most* or even what a useful definition of teacher quality might be. One reason for this difficulty is that teacher quality may need to be defined differently for different purposes. For example, the indicators of quality relevant to making initial hiring decisions may be different from the indicators used in granting tenure, rewarding excellent performance, or identifying and supporting struggling teachers. In addition to teacher contributions to student achievement, teacher quality may be evidenced by teachers who possess the following characteristics:

- Qualifications and experience appropriate to grade level and subject matter.
- High expectations for students, particularly those at risk for poor outcomes.
- Creation of a classroom environment that encourages all students to participate in worthwhile learning activities.
- Desire to help students achieve at high levels.
- Ability to motivate at-risk students to come to school and participate in class, even if their achievement scores do not show significant gains.
- Excellent skills in mentoring new teachers and acting as stabilizing forces in high-turnover schools.
- Willingness to work diligently with students with special needs, whose test scores may not reflect teacher contributions.

In this research synthesis, a one-size-fits-all definition of teacher quality is not used because a variety of occasions and purposes exist for which different definitions may be appropriate. Rather, this synthesis puts forth a framework for both conceptualizing and measuring teacher quality that allows for a number of interpretations. The framework includes two "inputs" (teacher qualifications and teacher characteristics), a process measure (teacher practices), and an outcome measure (teacher effectiveness). Within this framework, research conducted to measure teacher quality is examined and summarized.

How Can Teacher Quality Be Measured?

Although teacher quality has been operationalized using inputs, processes, and outcomes in a variety of studies, the outcome measure used for this research synthesis is student achievement on standardized tests. By limiting this synthesis to studies using standardized student achievement test scores as the outcome measure, it is possible to make some comparisons across studies so that a composite picture of the research emerges. For this reason, a number of studies were excluded from the synthesis because they used some other type of outcome (such as student grades, graduation rates, or student achievement on a local test rather than on a nationally normed test).

Measuring teacher quality using standardized achievement test scores is challenging for the following reasons:

- Standardized achievement tests were intended to measure student achievement and were *not* designed to measure teacher quality.
- It is difficult to sort out *teacher* effects (i.e., the contribution of teachers) from *classroom* effects (i.e., the contribution of peers, textbooks, materials, curriculum, classroom climate, and other factors).
- It is difficult to obtain linked student-teacher data that make it possible to connect specific teachers to student achievement test scores.

How Important Is Teacher Quality to Student Learning?

A great deal of research has been done on teacher quality using student learning as the outcome measure. Despite all the time and effort spent researching this topic, in only a few aspects of teacher quality does strong and consistent evidence suggest that certain dimension make a significant difference in student learning. Many aspects of teacher quality that have been measured have resulted in findings that are inconsistent across studies or have such small effects that they are of no *practical* significance, even when they are statistically significant. Much of the research currently being reported purports to provide evidence for the importance of some aspects of teacher quality; but when the studies are collected and synthesized, it becomes apparent that there is not a consistent message. Some studies report that a particular aspect matters, and other studies report that the same aspect of teacher quality does *not* matter.

Do Certain Aspects of Teacher Quality Have a Stronger Impact on Student Achievement for Specific Students, Subjects, or Grade Levels?

There is one aspect of teacher quality where a consensus across studies has clearly emerged: The effects of teachers with degrees in mathematics and appropriate certifications, and possibly higher level mathematics courses, appear to be strongly and consistently related to student achievement in mathematics. Although there is evidence for this result at both the elementary and the secondary levels, the findings are strongest at the secondary level, suggesting that such qualifications may be crucial for secondary teachers. Similar findings were *not* apparent for other subjects. This situation may be because fewer studies were found that focus on the impact of certification and other indicators of teacher quality on English, social studies, science, and other content areas. Or it may be that particular teacher qualifications simply do not matter as much in these other subjects. The reasons for this strong showing in mathematics will be further discussed in this research synthesis.

How Important Is Teacher Experience?

The synthesis of research in which teacher experience is used as an aspect of teacher quality suggests that experience matters, but it contributes differentially *only* in the first four or five years of teaching. During this time, teachers appear to gain in effectiveness (contribution to student achievement scores) but then they level off, which means that years of experience beyond the fifth year contribute little or no additional benefit in terms of student achievement. Experienced teachers may contribute to their schools in other important ways, however, including providing stability and serving as mentors to new or struggling teachers.

How Can Teacher Quality Be Better Understood?

As linked student-teacher data become more universally available and ways of measuring teacher effectiveness are refined through improvements in both the means of measuring teacher contributions to student learning and in the means of analyzing the resulting data, it should be possible to achieve greater consensus on defining teacher quality for various purposes (such as making hiring and tenure decisions, rewarding excellent teachers, and providing interventions for struggling teachers).

Using This Synthesis

This research synthesis provides an up-to-date, comprehensive compilation and review of the recent research regarding teacher impact on student achievement outcomes. Organized using a framework of inputs, processes, and outcomes, this synthesis is a “one-stop shop” for researchers and policymakers interested in the science behind claims about the link between teacher quality and student academic achievement. Although it is possible to locate references that support a particular viewpoint, good decisions are based upon good data, and this synthesis of high-quality studies is intended to support good decision making.

Introduction

What is the relationship between teacher quality and student achievement? What are the best ways to measure teacher quality? Is teacher *quality* the same as teacher *effectiveness*? How does teacher quality relate to the "highly qualified" teacher definition developed for the No Child Left Behind (NCLB) Act? The answers to these questions vary depending on which report, study, or policy brief is being examined. A clear consensus on the meaning of teacher quality has not yet been reached, although teacher quality is almost universally believed to be the most important school-based factor in student learning.

Many reports, studies, and research articles published in recent years suggest that teacher quality matters a great deal in terms of student learning. This research synthesis explores the evidence for this relationship in an effort to help identify which teacher qualifications and characteristics should be prioritized in educating and hiring those teachers who are most likely to have a positive impact on student learning. In addition, the framework developed for this research synthesis, when applied, will help put into perspective the many different aspects of teacher quality and how they have been measured.

The synthesis considers the various ways of defining teacher quality as well as the many ways it has been measured. The studies that are the focus of the synthesis use standardized student achievement test scores as outcome measures. The reason for focusing on teacher contributions to student achievement test scores is that this approach allows results to be compared across studies.

Other Research Syntheses on Teacher Quality

Before beginning this research synthesis, it is worthwhile to take a look at what other researchers found when they examined the literature on teacher quality and its relationship to student outcomes. As will be shown, this research synthesis differs from the others primarily in having a more comprehensive framework and a broader scope. Most other research syntheses in this area have focused chiefly on inputs and their relationship to student outcomes. In this synthesis, the framework will describe a wider range of qualifications, characteristics, practices, and outcomes (effectiveness) and analyze their relationships in ways that extend ideas about the how teacher quality is defined and measured.

Darling-Hammond and Youngs (2002)

Darling-Hammond and Youngs (2002) reviewed research on teaching qualifications and student achievement in order to counter arguments made in the U.S. Secretary of Education's annual report on teacher quality (Office of Postsecondary Education, 2002). Darling-Hammond and Youngs believed that the secretary was essentially calling for a lowering of standards for teacher qualifications.

Darling-Hammond and Youngs sought to *refute* the following specific assumptions: (1) teachers matter for student achievement, but teacher education and certification are not related to teacher effectiveness; (2) verbal ability and subject matter knowledge are the most important

components of teacher effectiveness; (3) teachers who have completed teacher education programs are academically weak and underprepared for their jobs; and (4) alternative certification programs have academically stronger recruits who are highly effective and have high rates of retention.

In their review of the empirical literature, Darling-Hammond and Youngs found little support for these four assumptions. They note that the second assumption is the most strongly supported because many research studies have found that verbal ability and subject matter knowledge are related to teacher effectiveness. They contend, however, that research also indicates that pedagogical coursework and student-teaching experience are at least as important in producing effective teachers. The fourth assumption also has some empirical support but only for select, well-designed alternative certification programs. Many other alternative certification programs produce markedly less-effective teachers who have high rates of attrition from the profession. Thus, the Darling-Hammond and Youngs review of the research confirms that some teacher qualifications may matter more than others but indicates that these qualifications often are mediated by the grade level and subject matter being taught.

Rice (2003)

Rice (2003) focused on five "teacher attributes": experience, preparation programs and degrees, certification, coursework, and teacher test scores. In discussing her findings, Rice makes a simple but important point: The findings "should be interpreted in the light of the availability of empirical evidence" (p. 48). She points out that a lack of evidence for a relationship between some attributes and student achievement may mean that empirical evidence was not readily available, rather than that no relationship exists. Rice (2003) concludes the following:

- Teacher experience matters, particularly in the first few years of teaching. More experience may be of greater importance for high school teachers than for teachers in earlier grades.
- Teacher preparation studies provide limited evidence of how teacher preparation programs improve teacher competency or student achievement. Program selectivity may be related to student achievement at the high school level, and high-poverty students may also gain more from teachers prepared in selective programs. Recent research on advanced degrees shows some evidence that such degrees may improve student achievement, but only in high school mathematics and to a lesser extent in high school science.
- Teacher certification seems to matter for high school mathematics, but there is little evidence of its relationship to student achievement in lower grades. There was no indication of a difference in student outcomes for teachers who gained certification through an alternate route.
- Teacher coursework, whether subject specific or in pedagogy, appears to have a positive impact on student learning at all grade levels, but subject-specific coursework matters most in secondary education. There may be a limit, however, to this positive effect: Requiring more courses for teachers does not translate into higher student achievement in a linear fashion.

- Tests that measure teacher literacy or verbal ability appear to correlate with both teacher performance and student outcomes. Although evidence for the impact of other types of test scores was mixed, there was an indication that teacher test scores are particularly important for the achievement of at-risk students.

Rice concludes from the available evidence that "more refined measures of what teachers know and can do (e.g., subject-specific credentials, special coursework taken) are better predictors of teacher and student performance than are more conventional measures (e.g., highest degree earned, undifferentiated course credits earned)" (p. 50). Rice's synthesis is a valuable contribution to the understanding about which qualifications matter most in terms of student achievement, but its scope is limited—primarily due to the lack of availability of empirical data on critical points.

Wayne and Youngs (2003)

Wayne and Youngs (2003) reviewed studies that related characteristics of teachers to student achievement. Their criteria for inclusion were stricter than those of most other research syntheses and are as follows:

- The collected data address teacher characteristics as well as the standardized test scores of the teachers' students.
- The data were collected in the United States.
- The design accounts for prior student achievement.
- The design accounts for student socioeconomic status.

Wayne and Youngs reported several interesting findings from their synthesis. Working with three studies, they found some evidence of a weak relationship between the selectivity (ranking) of teachers' undergraduate programs and student achievement. Most of the studies they examined (five out of seven) found that students benefited (in areas such as reading) from teachers with higher verbal scores. When degrees and coursework were examined, the authors found that all studies were positive concerning mathematics; that is, mathematics degrees and coursework appear to contribute to improved student achievement in mathematics. In addition, the authors reported that certification appears to matter only when the certification is in the subject area being taught and only for mathematics. It appears that the researchers were able to make stronger inferences about the importance of mathematics credentials to student mathematics achievement in part because there is a more substantial research base for this particular academic discipline.

Wilson and Floden (2003)

Wilson and Floden (2003) wrote an addendum to the report *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations* (Wilson, Floden, Ferrini-Munday, 2001). In this addendum, they synthesize research on teacher effectiveness in an effort to answer key questions, such as these: "To what extent does subject knowledge contribute to the effectiveness of a teacher? Is there a significant advantage to having an advanced degree in the subject taught as opposed to a subject major? To having a subject major as opposed to a minor?" Other

questions focus on pedagogical theory and knowledge, field-based experience, accreditation of teacher preparation programs, and alternative versus traditional preparation programs.

Although the Wilson and Floden addendum is focused on teacher effectiveness rather than teacher quality, the study is relevant because—for purposes of the synthesis—effectiveness is considered a component (and an outcome) of teacher quality. The particular sections focusing on the characteristics of new teachers that contribute to teaching effectiveness are of great interest because they answer some of the same questions that this research synthesis addresses. Relevant findings are summarized as follows:

- Findings on the impact of teachers' level of education were inconsistent, based on 14 studies.
- Findings on the relationship between teacher experience and student achievement were inconsistent, based on 12 studies.
- Findings on the relationship between teachers' verbal or general ability and student achievement were inconsistent, based on five studies.
- Findings on the relationship between teacher race, student race, and student achievement were inconsistent, based on six studies.
- Findings on the relationship between teachers' degrees and coursework and student achievement were inconsistent, based on 11 studies.
- Findings on the relationship between teacher preparation and student achievement were inconsistent, based on three studies.

Wilson and Floden summarized the evidence and discussed many of the problems they encountered in attempting to synthesize the research to answer the questions. One problem they noted is that many studies they examined did not tie teacher qualification and characteristics directly to student learning. Thus, although there was general agreement on what qualifications and characteristics were important, there was little evidence to support these assumptions. The authors also were relying on some older studies. Because linked teacher-student data have become more readily available and statistical software and methods have grown more sophisticated, more recent studies that will be discussed in this research synthesis add further nuance to the excellent work done by these authors.

Focus of This Research Synthesis

Although a number of research syntheses and reviews of the literature that focus on teacher quality have already been published, this synthesis adds to the existing literature in several ways. First, it focuses on the most recent studies, primarily those conducted since 2000. Second, it groups studies into a framework for evaluating teacher quality that may make it easier to talk about the components of teacher quality. Third, it provides summaries of the studies as well as tables that sort the studies in ways that should be useful when focusing on particular aspects of the studies. Fourth, it provides a summary table that presents thumbnail sketches of the studies examined for this synthesis.

Defining Teacher Quality

Teacher quality is a complex phenomenon for which no general and absolute agreement exists concerning an appropriate and comprehensive definition. One of the first dilemmas to resolve is the difference between *teacher* quality and *teaching* quality. *Teacher* quality implies that there is a set of inputs (such as certification, teacher test scores, and college degrees) that serves as indicators of who will be successful in the classroom. On the other hand, *teaching* quality implies that it is not what the teachers *have* in terms of training and certification, it is what they *do* in the classroom that indicates quality. Often, the two definitions are linked or even conflated, so that there is an assumption that *teacher* quality ensures *teaching* quality or that teaching *quality* is an outcome of teacher *quality*.

Perhaps more important, *teaching* quality can be broken down further into two dimensions: the task of teaching (what teachers do) and achievement (the student learning that teachers foster). Fenstermacher and Richardson (2005) elaborate on these concepts:

> Quality teaching could be understood as teaching that produces learning. In other words, there can indeed be a task sense of teaching, but any assertion that such teaching is quality teaching depends on students learning what the teacher is teaching. To keep these ideas clearly sorted, we label this sense of teaching *successful teaching*. (p. 186)

This viewpoint is useful for thinking about *teaching* quality, particularly successful teaching. Fenstermacher and Richardson's analysis is useful in this important respect: It clearly distinguishes what teachers *do* in classrooms from what students *learn* in classrooms. For purposes of the present research synthesis, student learning is the focus for both *teacher* quality and *teaching* quality. For that reason, all of the studies selected for examination as part of this synthesis have as an outcome standardized student achievement test scores. Although many other outcome measures could have been used, the one that is almost universally used and that allows for comparability among studies is the standardized test.

Framework for Teacher Quality

For this synthesis, examining recent studies and revisiting older studies led to the development of a new framework for determining teacher quality. The need for this framework stemmed from an effort to make sense of the many ways in which researchers have been measuring teacher quality over the years. This framework consists of four distinct but related ways of looking at teacher quality that are grouped into three categories, as follows:

Inputs

- Teacher qualifications
- Teacher characteristics

Processes

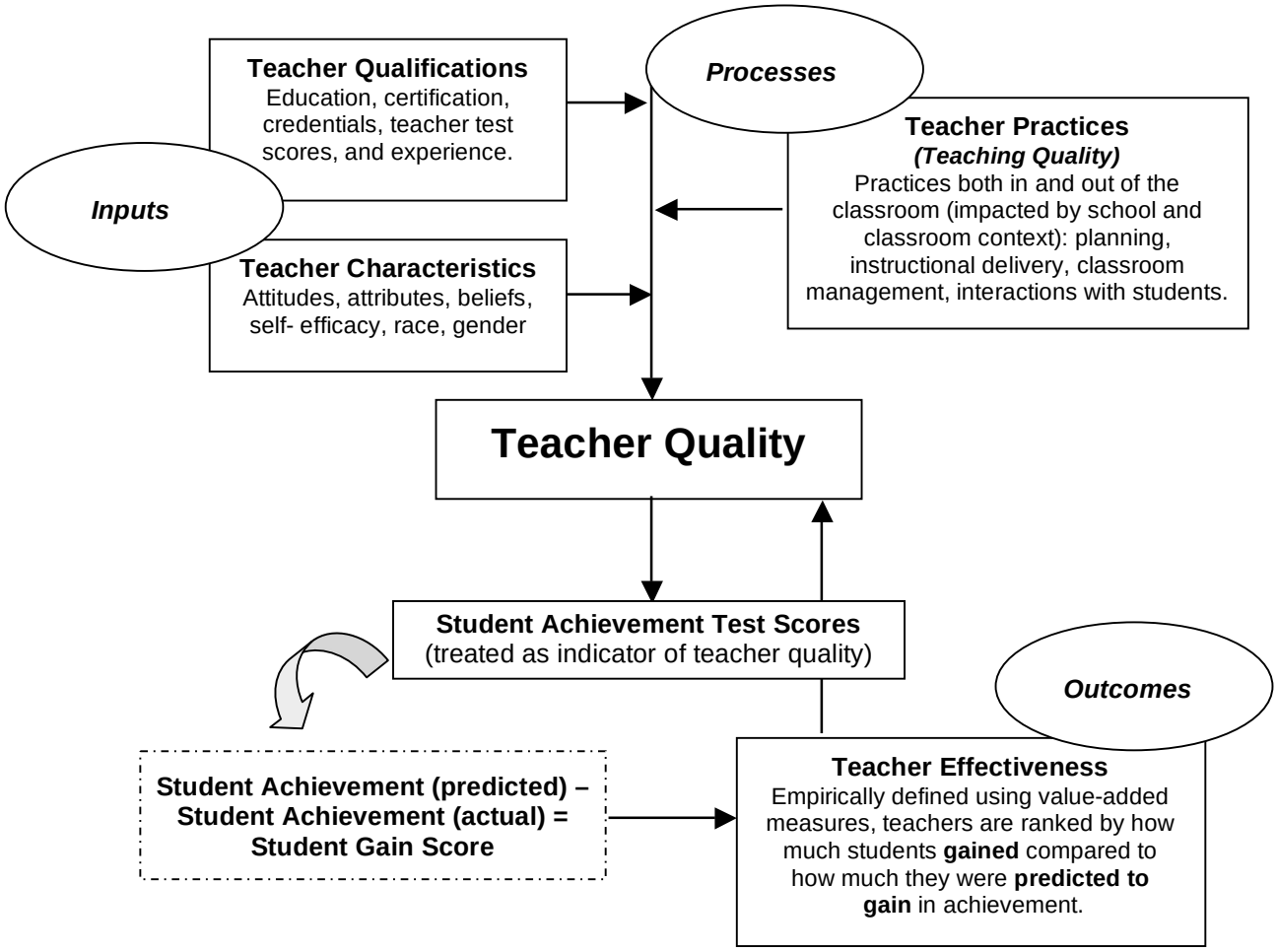
- Teacher practices

Outcomes

- Teacher effectiveness

Figure 1 shows how these four ways of looking at teacher quality are related.

Figure 1. Graphic Representation of a Framework for Teacher Quality



Note that teacher qualifications, characteristics, and practices are all used to define teacher quality and exist *independently* of student achievement, whereas teacher effectiveness is wholly dependent on student achievement. In other words, teacher effectiveness cannot be determined without outcomes such as standardized test scores. The other three ways of looking at teacher quality can be theoretically connected to student learning and measured with standardized test scores, but they exist whether or not they are measured. For example, teacher certification exists as a proxy for teacher quality, even if it is never connected to student outcomes. But teacher effectiveness exists only as a function of the link between teachers and their students' standardized test scores. What follows is a more in-depth description of each of these ways of looking at teacher quality.

Teacher Qualifications

The first strand of the framework for defining teacher quality focuses on teacher qualifications (also commonly called *teacher inputs*). Teachers' qualifications are among the resources they bring with them to the classroom and are considered important in establishing who should be allowed to teach. For determining teacher quality, however, the reliance on paper qualifications—for example, how many courses a teacher candidate took in a specific subject area or what score was received on a licensing test—may simply reflect the limitations of the data and research designs that are easiest to access. The reliance on paper qualifications as proxies for teacher quality seems to hold sway currently; thus, since the advent of NCLB, teacher quality often has been conflated with the idea of a highly qualified teacher.[1] Meeting NCLB requirements, of course, is no guarantee that teachers will be effective in their classrooms.

Qualifications also include teachers' coursework, grades, subject matter education, degrees, test scores, experience, certification, and credentials, as well as evidence of participation in continued learning such as internships, induction, supplemental training, and professional development. Experience also can be considered in this category because it is counted as a qualification for many purposes, including NCLB requirements.

The advantage of focusing on qualifications is that it allows education decision makers to use documents alone to estimate a teacher's potential effectiveness for licensing and hiring purposes, prior to any determination of the teacher's suitability for a position or effectiveness in the classroom. The major disadvantage of the qualifications definition of teacher quality is that a teacher can be deemed to be of high quality on paper yet perform poorly in the classroom.

Teacher Characteristics

The second strand of the framework for defining teacher quality focuses on teacher characteristics, including attributes and attitudes of teachers as well as immutable (or assigned) characteristics such as race and gender. Research in this area that links these characteristics to student outcomes is still relatively scarce. The advantage of viewing teacher characteristics in this way is that it expands the scope of teacher quality and thus creates an opportunity for greater precision in the definition of it. The main drawback to defining teacher quality in this way is that it focuses on characteristics that are often logically, ethically, or practically beyond the teacher's (or school's) ability to change.

Teacher Practices

The third strand of the framework for defining teacher quality focuses on examining teachers' actual classroom practices and correlating those practices with student learning outcomes. Evaluating teachers' questioning strategies and linking them to student learning is one example of a classroom practices mechanism. By this definition, teacher quality is ascertained not by the qualifications teachers have on paper but by what they actually do in the classroom with their

[1] For NCLB purposes, highly qualified teachers must possess the following inputs (paper qualifications): full state certification, bachelor's degree, and demonstrated subject matter competency in each of the academic subjects he or she teaches.

students, including instructional and classroom management practices, interactions with students, and performance of tasks. Higher correlations with what are considered better practices thus define good teaching. The focus, then, is not on assessing the connection between what individual teachers do but on correlating certain recommended practices and student outcomes.

The advantage to this definition over qualifications is that it focuses on the classroom, where the teacher and student interact and where learning actually takes place. The chief disadvantage of this definition is that evaluating teachers in their classrooms is difficult, time consuming, expensive, and subject to the complications of context (e.g., differences among urban and rural schools, high-poverty and wealthy schools, schools serving large numbers of English language learners, classrooms with students who have severe behavioral problems, and so on). Another disadvantage is that while researchers may focus on looking only at whether teachers are using one or two specific best practices, it is likely that teachers using these practices are also using other best practices. Thus, linking student learning outcomes to a handful of practices (and excluding all others) is virtually impossible. Similarly, another disadvantage is that studies examining teacher practices often do not control for other contributions to student learning (such as a classroom climate that is conducive to learning) or distractions that prevent students from learning (such as a disruptive classmate).

Teacher Effectiveness

A fourth strand of the framework for defining teacher quality is teacher effectiveness—as determined by growth in student learning, typically measured by standardized achievement tests. This strand most closely approximates a comprehensive measure of *teaching* quality rather than *teacher* quality because teacher effectiveness would be the empirical evidence that defines *teacher* quality and *teaching* quality, based on how much student learning a teacher fosters. Teachers might be considered high quality if their students learn significantly more than would have been predicted given those students' prior achievement.

Earlier Definitions of Teacher Effectiveness. It is worth noting that there has been a substantial shift during the past 30 years in how teacher *effectiveness* is defined and measured. At the 1978 Conference of the International Association for Educational Assessment, Schlusmans (1978) described the following eight ways of measuring teacher effectiveness:

- **Characteristics Deduced From a Theory.** Starting from existing educational, psychological or sociological theories, one deduces a number of characteristics of the effective teacher.
- **Characteristics Determined by the Pupils.** The evaluation by the pupils is used as the criterion for effectiveness.
- **Characteristics Defined by Specialists.** Inspectors and directors determine the characteristics of effective teachers from their own experiences with teachers and from their own theories.
- **Characteristics Derived From the Functional Analysis of the Teacher.** From the results of observation, surveys, and theory, a functional analysis of the teacher is made, on which conclusions about the characteristics of the effective teacher are based.

- **Characteristics Derived from a Role Analysis of the Teacher.** On the basis of a set of norms and expectations about teachers, characteristics of the effective teacher are formulated.
- **Characteristics Derived From Descriptive Research on the Teacher Population.** On the basis of characteristics discovered in the existing population of teachers, characteristics of the effective teacher are determined.
- **Empirical Research on Teacher Characteristics.** Characteristics of teachers, measured by observation scales and questionnaires, are tested for a specific criterion, such as the evaluation of teaching, the judgment of inspectors, the opinion of pupils and—in some exceptional cases—the achievement results of the pupils.
- **Predictive Research of Teacher Characteristics.** One tries to determine to what degree specific characteristics of trainees can predict a criterion of effectiveness, such as the obtaining of a diploma, the marks awarded, or the judgment of inspectors. (pp. 19–20)

Note that the seventh criterion describes measuring teacher effectiveness by the achievement results of the pupils "in some exceptional cases." Thus, teacher effectiveness was almost never measured through attempts to link student achievement with specific teachers or teacher characteristics. Now, however, using achievement results to determine teacher quality is becoming commonplace.

The Influence of Technology in Determining Teacher Effectiveness. Technology has made it much easier to connect teachers with data on student achievement. The proliferation of such data, the advent of powerful desktop computers, and advances in statistical software have made it possible to look at such linked data in new ways. Also, the movement toward school-level accountability for student achievement has provided the impetus to do so. In recent years, the focus has moved away from holding *schools* accountable for student achievement and toward holding *teachers* accountable. States and districts are increasingly experimenting with value-added models in an attempt to establish some measure of teacher effectiveness. In some states (such as Tennessee), this information has been used for research purposes—and, to a much lesser extent—as one of a number of factors a principal might use when evaluating teaching performance. In some districts (such as one in Houston, Texas), however, teachers are receiving substantial monetary rewards for improving student achievement. This situation has resulted in considerable turmoil, particularly when the rewards are accompanied by claims that the recipients are the very best teachers in the district (Associated Press, 2007). Not surprisingly, teachers who were not rewarded were left to wonder and grumble. Clearly, teacher effectiveness as measured empirically by student achievement has not yet reached a level of public approbation and acceptance.

Considering Teacher Quality Through a Lens of Teacher Effectiveness. Looking at teacher quality through an effectiveness lens means focusing on results that theoretically can be attributed to the other three strands of teacher quality (teacher qualifications, teacher characteristics, and teacher practices). However, it is impossible to determine from a value-added score which combination of qualifications, characteristics, and practices have contributed to student achievement. Using the effectiveness definition of teacher quality has the advantage of determining teacher quality without regard to teachers' paper qualifications, characteristics,

or practices. Thus, teachers who may not meet all (or any) of the qualifications for a particular position theoretically may still be deemed high quality if their students are performing better than expected.

On the other hand, a major disadvantage of the effectiveness definition is that it provides no mechanism for *predicting* high-quality teachers *prior* to their actual teaching. In other words, if teacher quality is to be determined solely by effectiveness, how can one decide who should be allowed to teach in the first place—before any student gains can be assessed? How can students be protected from ineffective teachers? This situation suggests that there is still a decided benefit to using assessments or other mechanisms that require prospective teachers to demonstrate a *minimum* level of competency before they are given teaching responsibilities. This concern is the fundamental rationale for the existence of teacher licensing requirements.

Use of the Teacher Quality Framework in This Research Synthesis

As has been shown, the four ways of looking at teacher quality—qualifications, characteristics, practices, and effectiveness—all have merit, but each also has drawbacks. In this research synthesis, the author will consider what the evidence says about teacher quality as determined by the four strands of the framework and attempt to put these findings into perspective.

Specifically, research that purports to measure teacher quality by linking paper qualifications (inputs), teacher practices, and characteristics with standardized tests scores will be examined. Research on the fourth strand, teacher effectiveness as measured by growth in student test scores, also will be examined. An attempt will be made to draw appropriate conclusions from the evidence that exists for these various types of measures, with the primary purpose of advancing efforts to ensure that all students—particularly special-needs and at-risk students—have an opportunity to learn at high levels. Clarifying exactly what constitutes teacher quality should help to further those efforts.

Brief summaries of the research in these four areas can be found in Appendix C. Its purpose is to guide those whose mission it is to understand teacher quality at a practical level, particularly those who educate teachers, hire teachers, or make policy decisions concerning teachers.

Before moving on to summaries of the evidence related to the four strands of teacher quality, it should be pointed out that there are many other ways to define and examine teacher quality. Most of the other options do not use student achievement as an outcome measure, however, and thus they do not fit into the framework created for this research synthesis. Another important lens through which to examine teacher quality is to evaluate the many ways that teachers contribute to their schools and thus to improving opportunities for teaching and learning throughout these schools. A teacher may take a leadership role in the school; apply for and win grants for educational innovations; serve as a mentor to new or struggling teachers; spearhead the implementation of reforms; develop ways to promote teacher collegiality; and work during the summer months to investigate the alignment between curriculum, materials, and tests.

One study that reviews a construct the authors call “collective teacher efficacy” illustrates an attempt to measure teacher quality as a group effect or aggregate of teachers within a particular

school, grade, or team, using student achievement as the outcome measure. Goddard, Hoy, and Hoy (2000) found a strong association, but not a causal relationship, between collective teacher efficacy and student achievement. Although teacher contributions such as these are crucial to the educational success of schools generally, there are few if any studies that link the contributions of *individual* teachers with student achievement as an outcome measure. Such alternative ways of looking at teacher quality are interesting and worthwhile; omitting them from the current research synthesis was a utilitarian decision and not meant to suggest that they are unimportant.

Using Student Test Scores to Determine Teacher Quality

Some important caveats should be considered before using standardized tests to measure teacher quality, either through connecting various inputs (qualifications and characteristics) or processes (practices) to students' scores, or through measuring teacher effectiveness using a value-added model.

First, the types of standardized tests given to students to measure achievement were never designed for the purpose of assessing teachers. They were not engineered to be particularly sensitive to small variations in instruction or to sort out teacher contributions to student learning from other factors that impact learning—school and classroom climates; peers; alignment among curriculum, standards, and tests; availability of materials that are aligned with what is tested; parental involvement in student learning; opportunities for teacher learning (such as high-quality professional development); and other factors.

Second, the use of value-added models for determining teacher effectiveness is controversial—and for good reason. Some researchers (particularly William Sanders, who designed and implemented value-added models for ranking Tennessee teachers) contend that because students' prior test scores are used as statistical controls in the formulas for calculating value added, there is no need to take into account other variables such as class composition and peer effects (W. L. Sanders & Horn, 1998). They believe that variables that might affect student test scores (such as poverty and school climate) are already included in the prior years' test scores, which are used to predict students' future achievement. However, many variables go into the making of a school or classroom within a school, and it is hard to imagine that teachers are solely responsible for students' test scores after controlling for students' prior achievement. In addition, some researchers (Braun, 2005; Kupermintz, 2003; Lockwood, Louis, & McCaffrey, 2002) are not convinced that the current generation of value-added models is sufficiently valid and reliable to use for evaluating individual teachers' effectiveness. In addition, using value-added models to rank teachers and determine teacher quality remains highly controversial. McCaffrey, Lockwood, Koretz, and Hamilton (2003) contend that major challenges to using value-added models for determining teacher effectiveness include incomplete data and confounding influences that impact student scores and that may not be included in the models (e.g., school effects).

Third, when state standardized tests are not aligned with standards, curriculum, and materials used in classrooms, student learning may not be reflected accurately in test scores. In other words, students may be learning plenty but what they are learning may not be what is being tested. States vary in the degree to which tests are aligned with standards, curriculum, and materials, which means that some states may have more accurate measures of teacher contributions to student learning than others. Because of this variability, teacher effectiveness as measured by student achievement tests may be useful for research purposes under some conditions, given that there are no other comparable outcome measures; however, it is more problematic when used for rewarding teachers or comparing teachers across states.

Methods

Because other useful research syntheses exist on the topic of teacher quality, it seemed most important for this one to focus on recent studies that may *not* have been included in previous research syntheses and literature reviews. Thus, most of the studies featured in this synthesis are from the past five or six years. However, a few older seminal studies have been included because they provide interesting approaches or findings and they also provide needed context for studies that came later.

Studies were selected based on the following criteria: (1) having as an outcome measure student achievement on a standardized, nationally normed test; and (2) having some type of measure of teacher quality, including those studies in which other factors contributing to student achievement also were measured. In general, the studies link individual student achievement to measures of individual teacher quality. However, some studies have aggregated or averaged data because linked student-teacher data were not available. Even now, obtaining such linked data persists as a serious challenge to doing research on teacher quality.

Studies were identified using Internet resources as well as library resources. Recommendations of suitable studies were solicited from experts, and research syntheses on teacher quality also were examined for possible leads. Given that the focus was primarily on the most current studies, the Internet proved to be the most useful source for identifying recently published studies.

Three appendixes provide at-a-glance overviews that may be helpful. Appendix A is an aid to understanding how the studies used different *types of variables* as indicators of teacher quality. Appendix B is a table describing the various *data sources* used in these studies. Finally, Appendix C is *summary table of studies*, included for quick reference, which highlights all of the studies summarized in this research synthesis and provides a brief description of the research.

A Note on Effect Size

In recent years, effect sizes have become increasingly important for describing research findings, for three reasons. First, effect sizes are useful in determining the actual measurable impact of an intervention (such as professional development or whole-school reform) or of a particular input into the classroom (such as teacher experience or certification). The effect size makes it easier to see just how important that intervention or input was in terms of student gains in test scores. Second, effect sizes allow for comparison across studies, meaning that studies reporting effect sizes can be evaluated to determine an average effect of a particular intervention or input or to attempt to rank the comparable worth of various interventions or inputs by larger or smaller effect size. Third, effect size estimates make it possible to compare findings from studies that used *different methods and measures*, making it possible to compare studies that used different methods of gathering and analyzing data.

In this research synthesis, effect sizes are reported in summaries of research findings only when the authors provide the effect size; no effect sizes were calculated independently for purposes of this research synthesis because of the difficulty of accurately calculating effect sizes in some types of studies, particularly where all statistics are not reported. In this research synthesis, all reported effect sizes are the authors' reports and have not been independently verified. This situation is important to clarify because, unfortunately, there is not a clear consensus in the research community about what an effect size represents and how it should be calculated and reported. In addition, the research base does not yet exist to help us interpret effect sizes across studies for most measures of teacher quality. There still are considerable gaps in the understanding of how to calculate effect sizes appropriately when looking at teacher effects, particularly taking into consideration multilevel models. (For more discussion on effect sizes and their calculation, see Appendix D.)

Research on Teacher Qualifications

Focus

This section focuses on the link between teachers' paper qualifications (including experience) and student learning—an area of continuing controversy, with some studies finding no or small effects of certification and experience and others finding significant positive effects.

Findings

Taking this group of studies as a whole, there appears to be strong consensus that mathematics certification matters, particularly at the secondary teaching level. However, evidence is lacking to support a similar relationship in other subjects. In addition, there is substantial evidence of yearly growth in teachers' ability as measured by their contribution to student learning in their first five years of experience. After the first five years, there is no evidence that increasing experience contributes additional impact.

Research Studies

Betts, Zau, and Rice (2003)

Betts, Zau, and Rice (2003) focused on the San Diego Unified School District for their research, which linked student and teacher data in elementary through high school, using 1998–2000 data. The population for this study consisted of teachers and students in 123 elementary schools, 24 middle schools, 17 high schools, and 5 charter schools. Many variables were included in their analyses, including school, student, and teacher characteristics. The authors used teachers' paper qualifications as teacher quality variables, including experience, level of education, credentials, and subject matter knowledge. They found that the correlations among these qualifications and student achievement varied substantially across grades and across subjects. According to their findings, elementary student gains in both mathematics and reading were higher when students were taught by an emergency credential teacher or a teacher with one year or less of experience, compared with a fully credentialed teachers with 10 or more years of experience—certainly a counterintuitive finding! Teachers with master's degrees contributed marginally more to increased mathematics scores than teachers with only bachelor's degrees. In middle school, gains in reading were correlated with teachers holding Ph.D.s in any subject (for English teachers). Students' scores in middle school and high school were negatively impacted by having a teacher who held only an emergency credential. In middle and high school mathematics, a teacher's mathematics authorization (a proxy for subject-area knowledge) was the best teacher-level predictor of student achievement.

Taken together, these results suggest that the contributions of various paper qualifications vary widely among subject areas and between grade levels. What matters in mathematics (subject knowledge) may not matter in reading, and what matters in the secondary grades (teacher credentials) may not matter in the primary grades.

Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005)

Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005) used teacher preparation as their measure of teacher quality in a study examining different pathways into teaching in New York City. The study used linked data for teachers and students in Grades 3–5 to examine differences in effectiveness among teachers entering the teaching force through traditional or alternative mechanisms. English and mathematics scores were used from New York's statewide tests, which are aligned to the state standards. More than a million student mathematics scores and more than 900,000 student English scores were used, along with data on more than 65,000 teachers.

The authors found differences in outcomes for teachers in their first year of teaching. For mathematics teachers, temporary license holders were found to be similar to Teaching Fellows,[2] while Teach for America teachers were similar to college-recommended (traditionally prepared) teachers in terms of their contribution to student achievement. For students' English achievement, the Teaching Fellows and Teach for America teachers performed worse than college-recommended teachers in terms of their contributions to student achievement. Temporary-license teachers fell between the college-recommended and alternatively prepared teachers. Teach for America and Teaching Fellows teachers' effectiveness in mathematics increased with time, however; second-year teachers from these pathways caught up to traditionally prepared teachers.

In the coming years, this ongoing study will provide a considerable amount of useful information about how qualifications (in this case, type of preparation) matter in terms of student achievement in both the short and long term. These preliminary results are interesting because they suggest that there are differences in teacher quality among teachers prepared in different ways. It is difficult to determine, however, whether these differences are due to the preparation and support (or lack of support) these teachers received or whether they are actually reflecting differences in backgrounds, aptitude, and characteristics of people who enter teaching from various alternative and traditional pathways. Additional research that may provide greater detail to help answer these questions in the future is currently underway.

Carr (2006)

Carr (2006) linked Ohio teachers' experience, degree level, and designation as highly qualified by NCLB requirements with student achievement as measured by Ohio's standardized proficiency tests. He used archival data from students and teachers in traditional and charter schools for the 2004–05 school year. Other variables linked by the author with student scores included student attendance, mobility, and disciplinary referrals. Controls included student socioeconomic status, learning disability status, race, and community type (urban versus nonurban). He also considered policy alternatives that could be tied to results, including increasing school funding, changing funding priorities, decreasing student-teacher ratios, increasing teacher quality, and improving student behavior.

[2] Teaching Fellows (www.nycteachingfellows.org) is a program designed specifically to help alleviate teacher shortages in New York City public schools. The program subsidizes candidates attaining master's degrees in shortage areas, particularly mathematics, the sciences, bilingual education, special education, Spanish as a foreign language, and English as a second language. The focus is on preparing teachers for high-needs school settings.

Carr's findings suggested that for public schools, teacher quality (i.e., highly qualified teacher status) was significant in 18 of 21 models but teacher experience and advanced degrees did not significantly contribute to student achievement (when controlling for highly qualified status). Teacher variables made no statistically significant contribution in charter schools. Although the teacher quality effects in public schools were statistically significant, they were not large. This finding suggests that NCLB-authorized paper qualifications alone account for only a small percentage of teacher contributions to student learning as measured by student achievement test scores.

Cavalluzzo (2004)

Cavalluzzo (2004) focused on certification from the National Board for Professional Teaching Standards (NBPTS) as a measure of teacher quality. She used linked student and teacher data on 108,000 student records in the Miami-Dade County Public Schools to examine the contribution of teachers' professional qualification in ninth- and tenth-grade mathematics. Teacher characteristics included in the model were experience, type of mathematics teaching certification, primary job assignment (mathematics or other), advanced degree, selectivity of undergraduate school, and whether the teacher had obtained National Board Certification. The study also controlled for a variety of student characteristics, including demographics; repeating grades; identification as gifted; suspension record; attendance; grade point average in core subjects; average teacher-assigned scores in mathematics for effort and for conduct, age, and grade level; whether the course taken was above or below the student's grade level; and enrollment in a limited-English-proficiency program.

The author found that with the exception of undergraduate school selectivity, each of the teacher quality indicators was significant and correctly signed in terms of contribution to student achievement. In addition, the author reported that compared with students whose teachers had never attempted National Board Certification, those students whose otherwise similar teachers passed the certification process had larger gains than those whose teachers had failed or withdrawn from the NBPTS accreditation process.

Besides the NBPTS findings, the author reported that having an in-subject-area teacher and regular state certification in high school mathematics were the greatest contributors to student achievement. However, the contributions of all of these qualifications were not of practical importance: Students with NBPTS teachers gained an average of 0.07 of a standard deviation after including school effects. Another factor that may have impacted her findings was that NBPTS teachers had better credentials overall than other teachers and were more likely to be teaching affluent, white, high-achieving, and gifted students.

Clotfelter, Ladd, and Vigdor (2006)

Clotfelter, Ladd, and Vigdor (2006) used linked data on nearly 4,000 North Carolina teachers and their fifth-grade students to determine the contribution to students' test scores of teacher experience, licensure test scores, advanced degrees, National Board Certification, and undergraduate institution attended. They determined that teacher experience had a significant positive effect on both reading and mathematics test scores and that teacher licensure test scores

had a statistically significant effect on mathematics scores; however, the regression coefficient of 0.012 with a standard error of 0.006 is of little practical importance. NBPTS status also had a statistically significant but not practically important effect on reading scores. In addition, the authors found a negative effect on student achievement for teachers with a master's degree, with a regression coefficient of –0.023 and a standard error of 0.012.

Although the authors confirm the contributions of a number of paper qualifications, their findings point to an issue that has appeared in many studies of this type: Results are different depending on the subject matter. Moreover, their results suggest that more teacher education does not necessarily result in improved performance for students. This finding calls into question the policy of many states of increasing the salaries of teachers who have or obtain advanced degrees.

Darling-Hammond (2000)

Darling-Hammond (2000) used National Assessment of Educational Progress (NAEP) reading and mathematics scores in her analysis of teacher qualifications and student achievement. She examined the correlation between the percentage of well-qualified teachers in the state and students' NAEP scores and determined that teacher qualifications are significantly and positively correlated with student achievement. Although the findings are interesting, there are a number of factors that should be considered in evaluating these results. Most importantly, there may be unknown differences among states that are tied to more rigorous requirements for teachers. For example, in states where a surplus of highly qualified teachers exists, the state can set high standards for teacher qualifications and still maintain a sufficient supply of teachers. Conversely, states that struggle to meet demands for teachers may lower teacher qualification requirements in order to ensure that classrooms are staffed. In addition, given the limitations of the data, it is not possible to make causal claims about the relationship between student achievement and teacher qualifications. Although correlation may be determined, it is possible that *both* variables (teacher qualifications and student achievement) are impacted by some other unknown variable.

Darling-Hammond, Holtzman, Gatlin, and Heilig (2005)

Darling-Hammond, Holtzman, Gatlin, and Heilig (2005) examined linked teacher and student data in Houston to determine whether teacher certification made a difference in student outcomes. Using a sample of 4,408 teachers in Grades 4 and 5 from the 1996–97 school year to the 2001–02 school year, they compared certified teachers with uncertified teachers (both Teach for America and non-Teach for America teachers). They found that uncertified teachers and those with the most nonstandard certifications had negative effects on student achievement gains, regardless of whether or not they were Teach for America teachers. Student achievement gains for Teach for America teachers were between one half to three months lower than student achievement gains for fully certified teachers, except in mathematics. However, the authors noted that Teach for America teachers who achieved full certification were about as effective as other fully certified teachers. For one group of alternatively certified teachers (those participating in a Houston-based certification program), students gained more on one reading test (the Aprenda reading test, a standardized Spanish language test). The authors theorize that because many of the

teachers in this program are Hispanic, the language match with Hispanic students may have contributed to achievement gains.

Decker, Mayer, and Glazerman (2004)

Decker, Mayer, and Glazerman (2004) evaluated the achievement of students who had Teach for America teachers compared with a control group of students with teachers who taught in the same grades in the same schools. Control teachers thus included traditionally certified, alternatively certified, and uncertified teachers. Using Grades 1–5 data from 17 schools, 100 classrooms, and nearly 2,000 students in Baltimore, Chicago, Los Angeles, Houston, New Orleans, and the Mississippi Delta during the 2002–03 school year, the authors found that Teach for America teachers had a positive impact on their students' mathematics achievement. The difference in growth was statistically and practically significant, with Teach for America teachers' students gaining about one additional month of mathematics instruction compared with control teachers. Furthermore, when comparing novice control teachers with Teach for America teachers, the differences were even more pronounced. However, there was no significant difference between student achievement in reading: Teach for America and control teachers contributed about equally to students' reading achievement. These results were stable across student subgroups, schools, and geographical regions.

Given that most Teach for America teachers are placed in hard-to-staff schools with high rates of teacher turnover, it is worth noting that these teachers fare as well as do scarce certified teachers. Thus, certification and teacher preparation may be less important as measures of teacher quality than the background characteristics of Teach for America teachers. It should be noted, however, that Teach for America teachers are an exceptional group (in terms of the selectivity of their undergraduate institutions and the rigorous screening process for acceptance) compared with other teachers entering the profession through an alternate route, and the findings may not hold for them.

Goe (2002)

Goe (2002) conducted a study on California schools that focused on examining aggregated school-level achievement as reflected in the state's Academic Performance Index for the school and a number of school-level student, school, and teacher variables. Using multiple regression, she found that two teacher quality factors showed small but significant *negative* correlations with student achievement: the percentage of emergency-permit teachers in the school and the percentage of first-year teachers in the school (controlling for credential status).

The study is hampered by the fact that it uses aggregated student and teacher data rather than linking individual student achievement scores with teachers. In addition, given that hard-to-staff schools typically have all three factors—low student achievement, many first-year teachers, and many uncertified teachers—it is possible that an unspecified (hidden) variable might explain the relationship. Thus, no causal claims can be made and the generalizability of the findings is limited by the study design.

Goldhaber and Anthony (2005)

Goldhaber and Anthony (2005) used North Carolina teacher data linked to student achievement scores to examine the relationship between National Board Certification and student achievement. Using elementary school records from 1996 to 1999, the authors matched 32,399 teachers to 609,160 reading students' test scores and 32,448 teachers to 611,517 mathematics test scores. The authors found statistically significant, but not practically important, student achievement gains for students whose teachers had completed National Board Certification (0.05 standard deviation in reading and 0.09 standard deviation in mathematics). However, they noted that students of these teachers were higher achieving and more affluent. The authors also noted that student achievement gains for teachers who would become National Board certified in the future (as determined with the longitudinal data) were just as effective as those who had already attained NBPTS certification.

Goldhaber and Brewer (1999)

Goldhaber and Brewer (1999) conducted a study examining teacher certification status and subject major and their relationships to student achievement using data from the National Educational Longitudinal Study of 1988. They found that students of teachers who had an undergraduate or graduate degree in mathematics performed better than students whose teachers did not have a mathematics degree (by 0.08 standard deviation, not of practical significance). In addition, they found that students of teachers with any type of certification to teach mathematics—including emergency, alternative, or standard certification—outperformed students whose teachers had no certification or who were certified in a subject other than mathematics. These results suggest that subject knowledge of mathematics may be more important than the type of certification in terms of the contribution to student achievement.

Hanushek, Kain, O'Brien, and Rivkin (2005)

Hanushek, Kain, O'Brien, and Rivkin (2005)[3] used teacher certification exam scores, educational attainment, teacher race, and years of experience to determine the links between these characteristics and student achievement in mathematics on the Texas Assessment of Academic Skills (TAAS). Data were archival records for school years 1989–90 through 2001–02 and included fourth- through eighth-grade students and teachers in one large urban district (about 230,000 student records). Using a value-added model, the authors found that experience predicted higher student achievement gains but only for the first few years of teaching. The authors determined that advanced degrees and certification exam scores were unrelated to student achievement scores on TAAS. In addition, they found that a match between student and teacher race improved achievement scores for minority students only. Moreover, they found that teachers who leave schools have significantly lower test score gains than those who stay in their placements.

[3] This study also could have been sorted into the Teacher Characteristics category because the authors examined teacher race and student achievement as well as paper qualifications.

Harbison and Hanushek, 1992

Harbison and Hanushek (1992) conducted a study using data from a Brazilian government-sponsored project. One focus of the research was an examination of whether resources would improve learning achievement. To ascertain this situation, they looked at teacher salary, education level, years of experience, participation in either of two inservice programs, and subject matter knowledge as determined by scores on the same tests (in Portuguese and mathematics) that were administered to their students. The study used a random sample of schools in the 218 high-poverty rural counties of northeast Brazil. The authors found that teacher education had a small, positive effect on second-grade mathematics students only. Teacher experience did not have a significant effect on students' test scores, and teacher participation in training did not contribute to improved student achievement.

Harris and Sass (2007)

Harris and Sass (2007) investigated the effects of teacher education and training using student, school, and teacher fixed effects. The authors used panel data on all public school students and teachers in Florida for two time periods (1995–96 and 2003–04), resulting in nearly 1 million matched student-teacher records in middle school alone. They found that preservice teacher training had little impact on student achievement. Further, they found that teachers' own test scores on the SAT verbal and quantitative sections had no impact on student achievement. Advanced degrees did not contribute to teachers' effectiveness and were even associated with reduced effectiveness in high school mathematics and middle school reading. However, the authors did find that that content-focused professional development seemed to make teachers more effective in middle and high school mathematics. Pedagogical content knowledge was positively associated with student test scores at the elementary and middle school levels but only in mathematics. In addition, there appeared to be a relationship between teacher experience and reading achievement in middle and elementary school students.

One particularly interesting finding from this study related to the impact of professional development on teacher effectiveness. According to the data, the effects from professional development participation were greatest three years after the professional development took place, meaning that it may take several years for the effects of such teaching learning experiences to have an impact on teaching. The authors found that content-oriented professional development had the strongest effect on student achievement.

Hill, Rowan, and Ball (2005)

Hill, Rowan, and Ball (2005) examined the effects of teachers' mathematical knowledge for teaching on first- and third-grade students' achievement, controlling for student and teacher covariates. In 115 schools, 699 teachers participated in the study and were followed over three years. A survey instrument designed by the authors to assess teachers' knowledge of teaching mathematics was used to score teachers. Using their instrument, which differentiated between pedagogy and mathematical knowledge for teaching, the authors determined that significantly better student results were linked to higher levels of teachers' mathematical knowledge. Of particular interest was that the scores on this instrument were better predictors of student

achievement than were teacher background variables such as preparation and certification or the length of time spent on teaching mathematics each day.

Kane, Rockoff, and Staiger (2006)

Kane, Rockoff, and Staiger (2006) estimated the effects of teacher certification status (certified, uncertified, and alternatively certified) as well as the effects of teacher education and experience on student achievement scores on the New York City standardized mathematics and readings tests for Grades 3–8. Their sample consisted of 9,849 mathematics and reading teachers matched to elementary and middle school students (95 percent match rate), excluding "mobile" teachers and those teaching high proportions of students with special needs. Using an educational production function, the researchers found that variations in teacher contributions to student scores: Students of internationally recruited teachers scored 0.02 standard deviations lower on math tests than students taught by regularly certified teachers, whereas Teach for America teachers scored 0.02 standard deviations higher on mathematics tests than regularly certified teachers. Students of New York City Teaching Fellows teachers scored 0.01 standard deviations lower than regularly certified teachers' students in reading. The authors also found that teacher effectiveness improved in the first years of teaching. The chief finding was that large within-group differences in effectiveness for each certification group surpassed the smaller between-group effects, meaning that the certification appeared to matter much less than other, unmeasured teacher characteristics independent of certification status.

McColsky, Stronge, Ward, Tucker, Howard, Lewis, and Hindman (2005)

McColsky et al. (2005)[4] examined the relationship between National Board Certification and student achievement. The study required several phases and was conducted on linked fifth-grade student and teacher data in three school districts in North Carolina. In the first phase, the research used two-level hierarchical linear modeling to develop effectiveness scores for each teacher, based on student test scores. In this phase, no significant differences were found between the aggregate student gains of NBPTS teachers and other teachers.

In the second phase, the researchers recruited the most and least effective teachers, based on their effectiveness scores and compared them to NBPTS teachers, using the following: (1) teachers' surveys of their own efficacy; (2) interviews about planning and assessment practices; (3) classroom observations focused on the level of cognitive demand of student and teacher questions, student behavior, and classroom management and intervention strategies; (4) analysis of the quality of reading comprehension assignments; and (5) teacher effectiveness ratings by trained classroom observers. For this phase, there were 25 NBPTS teachers and 282 non-NBPTS teachers. The researchers found that NBPTS teachers had slightly higher ratings on their planning practices and significantly higher ratings on the cognitive challenge of reading comprehension assignments. There were no significant differences in terms of the cognitive demands of student and teacher questions, classroom management strategies, or the numbers of disengaged or disruptive students. The most effective non-NBPTS teachers were rated significantly higher on the following four (of 15) teacher effectiveness dimensions than the least

[4] This study also could be sorted into the Teacher Practices category because it focuses on measuring specific teacher practices as well as paper qualifications (NBPTS certified).

effective non-NBPTS teachers: classroom management, classroom organization, positive relationships, and encouragement of responsibility.

In this study, it appears that what divides the most effective from the least effective non-NBPTS teachers is their ability to create and maintain a classroom climate conducive to learning rather than their use of specific instructional strategies.

Monk (1994)

Monk (1994) related the National Assessment of Educational Progress (NAEP) mathematics and science scores for three years with four teacher qualifications related to subject-matter expertise: mathematics and science coursework, major, degree, and experience. He found that teachers' subject matter expertise increases student learning gains but the benefit decreases after the fifth mathematics course. In science, student achievement was tied to teachers having taken at least four physical science courses or completing a science major. Mathematics pedagogy courses also were found to contribute to student achievement, as was the match of teachers' experience to the classes they taught. Teacher experience alone contributed to student achievement only for 11th graders. Teacher degree level was not significantly related to student achievement, with the exception that teacher degrees at the master's level and beyond appeared to be negatively related to student achievement.

Rockoff (2004)

Rockoff (2004) determined teacher quality by calculating the value added to student achievement in reading vocabulary, reading comprehension, mathematics computation, and mathematics concepts. Data from approximately 10,000 elementary students and 300 teachers in two New Jersey school districts were used in this study. The results of the analysis suggest that teacher fixed effects (teacher quality) have a small but significant effect on student achievement. Rockoff also found that teacher experience was positively related to student test scores in reading and mathematics but leveled off quickly in mathematics after the first two years of teaching.

Rowan, Correnti, and Miller (2002)

Rowan, Correnti, and Miller (2002)[5] sought to test various definitions of teacher quality against data from the Prospects National Longitudinal Study, a study mandated by Congress as part of the government evaluation of the Title I program. The authors used a three-level hierarchical linear growth model for each cohort of students (Grades 1–6) to examine "presage" variables (such as a certification status, advanced degrees, and experience) as well as "process" variables (such as use of active teaching methods and alignment of content coverage with assessments). The authors found the following effect sizes[6] for presage teacher quality variables: teaching experience and reading growth in Grades 1–3, d = 0 .07; experience and reading growth in Grades 3–6, d = 0.15; experience and mathematics growth in Grades 3–6, d = 0.18; and

[5] This study could be sorted into the Teacher Practices category as well, given the "process" variables described.

[6] A common way to express effect size is Cohen's d, in which the difference of the means of the treated and control groups is standardized by dividing by the pooled variance of the two groups. The resulting estimate expresses the magnitude of the effect in terms of standard deviations.

advanced degree in mathematics and mathematics growth for both cohorts, $d = -0.25$. They found the following effect sizes for process variables: time spent on whole-class instruction and reading growth, $d = 0.09$; time spent on whole-class instruction and mathematics growth, $d = 0.12$; alignment of content with assessments and reading growth, $d = 0.10$ to 0.18 for word analysis skills, reading comprehension, and writing process emphasis, respectively; and content alignment and mathematics growth in Grades 3–6, $d = 0.09$. Note that many of these effect sizes are too small to have practical significance.

Sanders, Ashton, and Wright (2005)

W. L. Sanders, Ashton, and Wright (2005) compared teachers with National Board Certification to other teachers, using more than 260,000 student records in mathematics and reading in two large North Carolina school districts. Of the more than 4,600 teachers included in the study, 281 were NBPTS mathematics teachers and 306 were NBPTS reading teachers. The authors tested four hierarchical models to examine student test data as a function of six fixed effects (year in school, previous year's test scores, race, sex, teacher experience, and NBPTS certification status) and a random teacher effect. They found that NBPTS teachers were not reliably more effective than the non-NBPTS teachers. In addition, they found that the variation among teachers with the same certification status was sufficiently large so that the small average differences between categories were trivial.

Vandevoort, Amrein-Beardsley, and Berliner (2004)

Vandevoort, Amrein-Beardsley, and Berliner (2004) also investigated the relationship between National Board Certification and student achievement. The authors administered surveys to teachers and principals and analyzed student tests scores for students in Grades 3–6 in 14 districts in Arizona. Thirty-five out of 80 (44 percent) of the NBPTS early childhood and middle childhood generalists agreed to participate. Test scores were collected for all students in the schools where these teachers taught. The authors found differential gains for students of NBPTS teachers equivalent to about 1.3 additional months of academic growth compared to students taught by non-NBPTS teachers.

The authors calculated pretest to posttest effect sizes independently for NBPTS-certified and non-NBPTS teachers and then converted the difference into grade equivalents using the work of Glass (2005), which found that an effect size of 1.0 is roughly equivalent to one year's academic growth on a standardized test. The authors estimate that an effect size of 0.10 is thus equal to one month of academic growth, and they report impacts of NBPTS-certified and non-NBPTS teachers accordingly. Differences in effect sizes between NBPTS-certified and non-NBPTS teachers ranged from 0.335 in third-grade reading to –0.230 (i.e., nothing) in fifth-grade mathematics.

Research on Teacher Characteristics

Focus

This category of looking at teacher quality focuses on characteristics such as (1) attitudes and beliefs, many of which are difficult to change; (2) immutable or assigned characteristics, such as race, ethnicity, and gender; and (3) characteristics that are potentially changeable, such as the ability to communicate in a second or third language.

Findings

There is not a clear consensus that any of the measured characteristics among these studies has an impact on student achievement. The data and research varied greatly. Some authors found significant relationships, but other authors researching the same characteristic did not find evidence for these relationships. Clearly, there is much more research to be done in this area.

Research Studies

Dee (2004)

Dee (2004) compared the achievement of students assigned to teachers of the same race with similar students who were assigned to teachers of a different race. All of the students were compared with those in the same grade and in the same school who were randomly assigned to teachers' classrooms. The authors contrasted same-race achievement results with different-race achievement results, using data from Tennessee's Project STAR (Student Teacher Achievement Ratio) class-size experiment. There were 23,883 cases for mathematics and 23,544 cases for reading, linked to teachers. The author found that, for black children, having a black teacher for one year was correlated with 3–5 percentile point increases in mathematics achievement. Similarly, reading scores for black pupils with a black teacher were 3–6 percentile points higher. White students placed with a white teacher scored 4–5 percentile points higher in mathematics, whereas the difference in reading was mixed by gender, with boys scoring 2–6 points higher and girls scoring about the same. Thus, students in the same teachers' classroom could have somewhat different educational outcomes based on whether they were the same race as the teacher.

This study is interesting because of the experimental design: Random assignment of students means that the possibility of students being sorted along race or other characteristics was avoided. This design lends additional credence to the findings, supporting Dee's contention that more efforts need to be made to recruit black teachers. Dee noted that the positive effects of being assigned to a teacher of the same race appeared to be cumulative, and he suggests that three to four consecutive years with a same-race teacher might contribute to closing the achievement gaps for black students.

Ehrenberg, Goldhaber, and Brewer (1995)

Ehrenberg, Goldhaber, and Brewer (1995) examined data from the National Education Longitudinal Study of 1988 (NELS 88) to determine whether teacher race, gender, and ethnicity matter in terms of student achievement. The authors found little evidence of an association with any of these teacher characteristics and student achievement, but they did find interesting evidence that teachers may have evaluated their students differently based on gender. Specifically, they found that in mathematics and science, white female teachers evaluated white female students more favorably than did white male teachers. This finding suggests better rapport among some combinations of teachers and students but does not directly support differential effects on student achievement.

Goddard, Hoy, and Hoy (2000)

Goddard, Hoy, and Hoy (2000) focused their analysis on collective efficacy among teachers, measured by assessing group competence and task analysis orientations, aggregated to the school level. They linked these scores to aggregated student achievement scores. Forty-seven randomly selected schools in a Midwestern urban school district provided names of faculty members to be surveyed (452 teachers responded). Achievement data on 7,016 students taught by these teachers also were obtained from the district. Using hierarchical linear modeling with student race, gender, socioeconomic status, and school size as covariates, the authors found a significant association between collective teacher efficacy and student achievement.

Because data for this study were collected by surveying a small number of teachers in each school, and because both teacher and student data were aggregated to the school level, it is difficult to determine whether the authors' claims of an association are justified. There may be other excluded variables at work that cause *both* the increased sense of efficacy and better student achievement.

This study also is instructive as one of the more recent examinations of the impact of teacher efficacy on student achievement. This construct, and the related construct of teacher expectations, appears to be the chief teacher personality characteristic that has been associated with student achievement in the empirical literature (e.g., Armor et al., 1976; Ashton & Webb, 1986; Moore & Esselman, 1992; Ross, 1992). Although teacher dispositions or personality characteristics may contribute to effectiveness, there does not appear to be any research that directly investigates the relationship between teacher personality characteristics (e.g., efficacy, authority, management style, persistence, and positive and negative feelings) and student achievement on standardized tests. Thus, the relationship between teacher personality characteristics and student achievement lacks an empirical research base.

Leana and Pil (2006)

Leana and Pil (2006) focused on examining social capital as operationally defined in a survey constructed by the authors. Survey items assessed teachers' information sharing, trust, and shared vision. The quality of teachers' instruction was rated through a survey in which parents reported their satisfaction with teaching methods, materials, and opportunities to learn. The

authors also included years of teaching experience in their analysis. Student achievement was measured using state standardized tests on mathematics and reading achievement, aggregated to the school level. They used a variety of qualitative data-collection strategies, including teacher and administrator interviews, observations of school processes and instructional quality, focus groups, teacher surveys (80 percent response rate), principal "diaries" (93 percent response rate), and archival data on student test scores. Using data from 88 out of 95 schools (elementary and secondary) in an urban Northeastern district, the authors performed a regression analysis with average mathematics and reading scores as the dependent variable and student poverty and average teacher experience as covariates.

The authors found that internal social capital—defined as teachers' information sharing, trust, and shared vision in a collaborative professional community—was significantly associated with both parental satisfaction with the quality of instruction and student achievement in mathematics and reading. They also found that instructional quality appeared to mediate the relationship between internal social capital and mathematics achievement. In reading, instructional quality predicted achievement but did not appear to mediate the relationship between internal social capital and reading achievement.

This study is particularly interesting because it takes a sociological perspective on school interactions and uses both qualitative and quantitative methods to develop the argument that how teachers relate to each other collaboratively—not just what they do instructionally—is important for student achievement,.

Research on Teacher Practices

Focus

The focus of this category of looking at teacher quality is on the connection between what teachers do in their classrooms—teaching practices and behaviors—and student learning. Many of these studies used observation protocols to document and evaluate what teachers did with their students.

Findings

Although most of the studies summarized in this section found some positive correlation between what teachers practice and student achievement, the results generally are not statistically or practically significant. In addition, a number of the studies have questionable research designs or use data, methods, or instruments that may not be appropriate to the goals of the research. Thus, there is an overall lack of findings that are both strong (i.e., significant) and convincing (i.e., appropriate design, methods, and instrumentation).

Research Studies

Borman and Kimball (2005)

Borman and Kimball (2005) adapted a standards-based teacher evaluation system from Charlotte Danielson's (1996) *Enhancing Professional Practice: A Framework for Teaching*[7] and used it to correlate teachers' scores with student achievement. They used data from 131 Grade 4 teachers linked with 2,527 students, 135 Grade 5 teachers linked with 2,176 students, and 131 Grade 6 teachers linked with 2,632 students in a Nevada school district. Teacher experience was included as a covariate. Hierarchical linear modeling was used to estimate teacher effects on classroom mean achievement. The authors increased sample size by using only Domain 1 (planning and preparation) and Domain 3 (instruction) of the evaluation system because only probationary teachers had evaluations on all four domains.

The authors found that teacher quality as determined by standards-based evaluation contributed slightly to student achievement. Teachers in the 84th percentile and above taught students whose average achievement was one tenth of a standard deviation higher than that of students of teachers in the 16th percentile and below. One possible confounding variable in the study was

[7] Written by Charlotte Danielson, the *Framework for Teaching* is founded on a research base developed by Carol Dwyer (1994) for the creation of the ETS Praxis III. The *Framework for Teaching* was explicitly created to provide a mechanism for assessing experienced teachers and is aligned with accepted standards for teaching, including those of Interstate New Teacher Assessment and Support Consortium (INTASC) and NBPTS. The *Framework for Teaching* defines 22 components of practice within four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities. The developmental stages of each component are articulated across four levels of performance that illustrate unsatisfactory, basic, proficient, and distinguished practice.

that teachers of less advantaged students may be unfairly evaluated as being less effective.[8] Given the challenging conditions in many high-poverty schools, this situation is an important consideration.

The very small differences found among the "best" and "worst" teachers in the sample might be interpreted in a number of ways:

- It is possible that the evaluation framework was not sensitive enough to pick up key differences in teaching practices, at least when limited to only two of the four domains.
- The student assessment may not have been sensitive to instructional differences; that is, the differences in instruction may not have greatly influenced students' responses on the tests.
- There may be other contributors to students' scores that were not measured by the evaluation instrument and which would account for additional variance among students' scores.

Cohen and Hill (1998)

D. K. Cohen and Hill (1998) used teachers' self-reported instructional practices through a 14-item survey consisting of questions about conventional practices and practices relating to the 1985 *Mathematics Framework for California Public Schools: Kindergarten Through Grade Twelve*. (California Department of Education, 1985) to determine their impact on students' mathematics scores in California. Of particular importance in this study is that the test used to measure student achievement was the California Learning Assessment System (CLAS) mathematics test, which was aligned to the 1985 *Mathematics Framework*. The study focused on determining whether the higher level of usage of *Mathematics Framework* practices was related to improved student achievement on CLAS. The results of the study suggested that there was a modest relationship between using *Mathematics Framework* practices and student scores on CLAS. Moreover, teachers' attendance at curriculum workshops, use of replacement units, and learning about CLAS also were related to higher CLAS scores.

The findings from this study provide important evidence on two fronts. First, the findings provide evidence that what teachers do instructionally matters. Second, the findings indicate that teachers' participation in professional development activities designed to change instructional practice may impact student achievement. There has been little research that provides evidence of a link between professional development and student learning, so this is a particularly important finding.

Limitations of the study included design and methodological issues, such as self-reports of teacher practices, use of absolute rather than gain scores, questions about the technical quality of the CLAS test, attrition among study participants, and difficulty in interpretation because of the effects of aggregating data. However, this study remains important because of the large sample

[8] This concern is substantiated by Jacob and Lefgren's (2005) finding that administrators appeared to discriminate against untenured teachers in their evaluations.

size and the direct links among professional development, teacher practices, and student outcomes that were studied.

Frome, Lasater, and Cooney (2005)

Frome, Lasater, and Cooney (2005)[9] used information on teacher characteristics for middle school teachers linked with eighth graders' achievement test scores in Georgia. Using data on teacher experience and education along with results from a survey administered to eighth graders, they found that of 11 teacher quality measures, the following four were significantly and positively related to student achievement:

- **Teacher Motivation and Expectations for Students.** Higher (student) ratings for motivation and expectations correlated with higher achievement.
- **Instructional Practices.** Higher (student) ratings for practices considered to be effective by the researchers were correlated with higher student achievement. Practices included group work on challenging assignments, oral presentations and written reports on mathematics projects, and explanations of solutions to the class.
- **Mentoring/Induction Experiences.** The percentage of teachers within a school who participated in mentoring/induction was significantly and positively correlated with students' mathematics achievement scores.
- **Content and Pedagogical Coursework.** The percentage of teachers within a school with a major in mathematics education was significantly correlated with students' mathematics achievement scores.

Although this study utilized an interesting source of evidence—student surveys combined with paper qualifications for teachers—it is limited in generalizability because results are aggregated to the school level, rather than linking individual teachers with their students' own survey ratings.

Gallagher (2004)

Gallagher (2004) conducted an in-depth, mixed-methods study of one Los Angeles elementary charter school serving approximately 1,200 students. Thirty-four teachers were evaluated on three occasions during the school year across 10 domains, including lesson planning, classroom management, special education inclusion, technology, and subject-specific areas. The evaluation rubric was based on Danielson's (1996) *Framework for Teaching*. These scores were then linked with students' value-added scores (growth compared to predicted growth). The author found that there were significant differences in student achievement relative to teachers' evaluation scores. In particular, literacy and composite evaluation scores were significantly related to student achievement, whereas mathematics and language arts scores were not. Gallagher also correlated teacher certification and experience data with student achievement and found no relationship with student test scores.

[9] This study also may fall under the categories of Teacher Qualifications and Teacher Characteristics.

In the qualitative component of the study, the author analyzed documents and conducted interviews with 12 teachers and three evaluators to try to understand the differences between evaluation scores and effects in reading and mathematics. Based on these interviews, he concluded that alignment and consistency in the pedagogical approach were factors in the correlation between literacy evaluation scores and student achievement.

Heneman, Milanowski, Kimball, and Odden (2006)

Heneman, Milanowski, Kimball, and Odden (2006) used a standards-based evaluation system to conduct a multiyear mixed-methods study investigating the validity of teacher evaluation systems. They worked with four sites throughout the country: Cincinnati, Ohio; Los Angeles, California; Reno/Sparks, Nevada; and Coventry, Rhode Island. The evaluation instruments were modifications of Danielson's (1996) *Framework for Teaching* and encompassed all four of its domains: planning and preparation, the classroom environment, instruction, and professional responsibilities.

Using linked student and teacher data, the authors assessed the relationship between student achievement and teachers' performance evaluation scores. They used a value-added model in which achievement was estimated based on prior achievement and other student characteristics. The authors found positive relationships between teacher evaluation scores and student achievement gains, although there was considerable variability across sites. In the Vaughn Charter School (Los Angeles), the correlation over three years averaged 0.37 in reading and 0.26 in mathematics. In Cincinnati, the correlation averaged 0.35 in reading and 0.32 in mathematics. Smaller correlations were found at the other two sites, with averages of 0.22 for reading and 0.21 for mathematics in Reno/Sparks (Nevada) and 0.23 for reading and 0.11 for mathematics in Coventry (Rhode Island).

Although the goal of the study was focused on the evaluation instruments themselves, it is worth noting that there was a fairly high correlation (at least in two of the schools) between what the teachers were observed to be doing in their classrooms and their students' achievement gains.

The authors speculated that the higher correlations in two of the sites were likely due to using multiple evaluators and, in Cincinnati, highly trained evaluators. Moreover, teachers at Vaughn Charter School had a shared understanding of what constituted good teaching. At the sites with lower correlations, the evaluations were conducted by a single evaluator who had less training.

Holtzapple (2003)

Holtzapple (2003) used a standards-based teacher evaluation system based on Danielson's (1996) *Framework for Teaching* to compare student achievement with teachers' evaluation scores. Focusing on 246 comprehensively evaluated Cincinnati Public School teachers in Grades 3–8 in 2000–02, the author examined the achievement of students linked to the teachers in the study using a value-added model of predicted achievement versus actual achievement. The author found that teachers who received low ratings on the instructional domain of the teacher evaluation system had students with lower achievement scores than would have been predicted by prior achievement. She also found that teachers with "advanced" or "distinguished" rankings

on this instrument generally had students with higher than expected test scores, whereas teachers rated "proficient" had students with average gains.

Jacob and Lefgren (2005)

Jacob and Lefgren (2005)[10] compared subjective principal assessments of 202 teachers with paper qualifications such as education and experience, and they linked these ratings and qualifications to value-added student scores. They found that the principals' assessments of teacher effectiveness were significantly better at predicting student achievement (based on a predicted score) than teacher experience or education, particularly in mathematics. However, the researchers also found that principals rated male teachers and untenured teachers lower than would be expected given their students' achievement gains; that is, those teachers were often more effective than the principals' evaluations would have suggested.

Kannapel and Clements (2005)

Kannapel and Clements (2005) conducted research designed to determine what made high-performing, high-poverty schools different from other high-poverty schools. They examined 26 high-poverty elementary schools in Kentucky using a standardized school audit instrument developed by the state. They selected eight of these schools based on high ratings on the audit. When these schools were compared with low-performing, high-poverty schools, differences were noted in a number of areas. In terms of teacher quality, the authors reported that teachers in the high-performing, high-poverty schools were more likely to conduct frequent assessments and offer students feedback; deliver instruction aligned to learning goals, assessments, and diverse learning styles; demonstrate high expectations for student performance; participate in collaborative decision making and ongoing, job-embedded professional development; and use student achievement data for staff development purposes.

Kimball, White, Milanowski, and Borman (2004)

Kimball, White, Milanowski, and Borman (2004)[11] examined the relationship between teacher evaluation scores and student achievement in nine grade-test combinations in Washoe County. The evaluation system used was adapted from Danielson's (1996) *Framework for Teaching* and rated teachers on the following: (1) pedagogical and content knowledge, (2) coherent lesson design and sequencing that correspond to student assessment, (3) adaptability to meet student learning needs, and (4) ability to engage students cognitively with strategies appropriate to learning goals. The teachers included 123 third-grade teachers, 87 fourth-grade teachers, and 188 fifth-grade teachers. Data included about 43 percent to 45 percent of all students in the district with pretest and posttest scores, and 50 percent to 70 percent of all evaluated district teachers who could be linked to qualifying students. Using a two-level hierarchical linear model, the authors estimated teacher effects on student achievement after regressing out student demographic characteristics and pretest scores.

[10] This study also could be sorted into the Teacher Qualifications category because it focuses on paper qualifications as well as teacher practices.

[11] This study also could be sorted into Teacher Qualifications category because paper qualifications are considered.

The authors found that teacher practices, as measured by the evaluation instrument, contributed slightly to student achievement. Although all of the correlations were positive, only two grade-test combinations were statistically significant. The authors also concluded that evaluation scores were stronger predictors of student achievement than were teacher education and experience.

Marcoulides, Heck, and Papanastasiou (2005)

Marcoulides, Heck, and Papanastasiou (2005) examined student perceptions of school culture and related them to student achievement. The authors used data from 1,026 eighth-grade students in secondary schools in Cyprus, which was collected as part of the Third International Mathematics and Science Study (TIMSS). As part of the assessment, students completed a survey in which a number of questions were asked about their teachers' strategies and practices used to help students learn. Students answered questions about the extent to which they worked on projects, discussed practical problems, and worked on problems relevant to their everyday life, as well as the extent to which teachers checked and discussed homework assignments, and aligned assessment and curricular practices. The authors found a 0.32 correlation between student perceptions of classroom practices and achievement. Interpretation of these findings is somewhat hampered by the lack of controls for student and school prior performance.

Matsumura, Garnier, Pascal, and Valdés (2002)

Matsumura, Garnier, Pascal, and Valdés (2002) reported on the technical quality of a measure examining the quality of classroom assignments. Developed as part of Los Angeles Unified School District's (LAUSD) proposed accountability system, the instrument was tested on 181 teachers randomly selected from 35 LAUSD schools. Fifty teachers submitted three language arts assignments, including four student work samples. Raters then scored the assignments along a number of dimensions (cognitive challenge, clarity of goals, clarity of grading criteria, and overall quality). Each submission was scored by five raters using a rubric. The researchers then performed hierarchical linear modeling with scores from the Stanford Achievement Test, 9th edition (SAT-9) as the dependent variable and effects estimated at the teacher level. They found that the quality of secondary teacher assignments as measured by their instrument predicted 0.08 of the variance in language arts achievement scores.

Matsumura, Slater, Junker, Peterson, Boston, Steele, and Resnick (2006)

Matsumura et al. (2006) conducted a pilot study of the Instructional Quality Assessment (IQA) toolkit in five urban middle schools. The IQA provides protocols to rate teachers' instruction through observations as well as through an analysis of the teachers' assignments and related student work samples. The authors found that the quality of instruction was highly variable within schools. They also determined that teacher observations were significantly related to scores on ratings of assignments in mathematics but not in reading. The relationship between the IQA and student achievement also was examined using linear regression. After controlling for students' prior achievement, ethnicity, socioeconomic status, language, and individualized education program status, the IQA rating predicted several reading and vocabulary subscores on the Stanford Achievement Test, 10th edition (SAT-10)). However, only the observation component of the IQA, the procedures subscore, predicted mathematics achievement.

The authors' findings suggest that evaluations of teacher quality (through observations or through examining lessons and work samples) are more or less valid, depending on the subject matter. It is possible that there are differences in mathematics and reading instruction that are not accounted for in the design of the instrument. However, there are some correlations between the instrument ratings and student outcomes, suggesting that the researchers are on the right track.

McCaffrey, Hamilton, Stecher, Klein, Bugliari, and Robyn (2001)

McCaffrey, Hamilton, Stecher, Klein, Bugliari, and Robyn (2001) used a teacher questionnaire to correlate self-reported instructional practices with student achievement scores among 10th-grade mathematics students. In a large, urban school district, 220 of 225 teachers of 10th-grade mathematics students returned questionnaires. The authors designed the questionnaire to assess teachers' use of instructional practices aligned to the standards of the National Council of Teachers of Mathematics (NCTM).[12] Using ordinary least-squares regression, the authors determined that more frequent use of practices aligned to the NCTM standards was associated with higher test scores among students in integrated mathematics courses—that is, courses designed to be consistent with the reforms recommended by the NCTM. However, there was no significant relationship between greater use of NCTM practices and mathematics achievement in other courses.

Milanowski (2004)

Milanowski (2004) analyzed the relationship between teacher evaluation scores and student achievement in a large Midwestern district using value-added measures. He used an evaluation system based on Danielson's (1996) *Framework for Teaching*, with 212 teachers in Grades 3–8 in Cleveland. He found small to moderate correlations between teacher evaluation scores and student growth. The average correlations were 0.27 in science, 0.32 in reading, and 0.43 in mathematics.

Newmann, Bryk, and Nagaoka (2001)

Newmann, Bryk, and Nagaoka (2001) examined teacher quality by looking at the intellectual demands of assignments given to students. They scored each assignment by the degree to which it required the construction of knowledge (through disciplined inquiry) in a way that gave it value beyond classroom learning. The authors collected 2,017 assignments—rated as either typical or challenging—from third-, sixth- and eighth-grade Chicago teachers. Trained scorers rated the intellectual demands of the assignments. These scores were subsequently matched to student achievement scores and analyzed using a three-level hierarchical linear model. Covariates used in the analyses were prior-year test scores, as well as student race, gender, and socioeconomic status.

The authors determined that in classrooms with high intellectual-demand assignments, students' learning gains were 20 percent higher than the national average for the Iowa Test of Basic Skills. In classrooms with low intellectual-demand assignments, the gains were 22 percent to 25 percent

[12] For a complete list of the NCTM standards, visit http://standards.nctm.org/document/appendix/numb.htm.

lower than the average in reading and in mathematics. The impact on the Illinois Goals Assessment Program (IGAP) was even greater, with high-demand assignments adding "standard effect sizes" of 0.43, 0.64, and 0.52 for the reading, mathematics, and writing portions, respectively. Use of high-demand assignments was not related to student demographics and benefited both high- and low-achieving students. However, high-achieving students benefited more from high-demand assignments in reading whereas low-achieving students benefited more from high-demand assignments in mathematics.

Rowan, Chiang, and Miller (1997)

Rowan, Chiang, and Miller (1997)[13] applied a general model about employee performance to the National Education Longitudinal Study of 1988 (NELS 88)to explain the effects of teachers on student achievement in 10th-grade mathematics. They focused on the variables of teacher ability, motivation, and work situation. The authors operationalized these variables by focusing on NELS 88 items that related to teachers' subject-matter knowledge, use of higher-order thinking instructional strategies, self-efficacy as applied to teacher expectations for student outcomes, and whether teachers worked in a restructured school environment. The authors found small effects on mathematics achievement related to teachers' subject-matter knowledge, their expectations for student outcomes, and their placement in school environments with shared decision making and common planning periods.

Schacter and Thum (2004)

Schacter and Thum (2004) examined 12 dimensions of teacher practices to try to determine between-teacher variation in student achievement gains. The 12 dimensions are teacher content knowledge, clarity of lesson objectives, presentation, lesson structure and pacing, relevance and challenge of activities, questioning skill, feedback, effective use of grouping, encouragement of thinking, motivation, environment, and teacher knowledge of students. The teachers were rated on these dimensions by trained graduate students using rubrics during two scheduled and six unscheduled visits during the course of a school year. Fifty-two elementary school teachers in Arizona volunteered to participate, and student achievement data were collected from their students in Grades 3–6 for reading, mathematics, and language arts. They used a mixed statistical model to determine the relationship between teachers' rating and student gains. They found that 84 percent of variation among teachers could be accounted for by the ratings on these dimensions.

Smith, Lee, and Newmann (2001)

Smith, Lee, and Newmann (2001) examined instructional approaches used by Chicago elementary teachers and their relationship to student achievement in mathematics and reading on the Iowa Test of Basic Skills. Data from more than 5,500 teachers and 110,000 students were used. The authors used hierarchical linear modeling, controlling for student gender, race, poverty, retention history, grade level, average ability level, problem behaviors, attendance, average income for students' parents, average achievement, racial composition, school instability, and school size. They found that didactic instruction is more common in higher

[13] This study could be sorted with the Teacher Qualifications category as well.

grades, low-level classes, "problem" classes, large schools, low-income schools, schools with low prior achievement, and predominantly African-American schools. They determined that the use of didactic versus interactive teacher methods was related to negative achievement growth (0.04 below average) and higher levels of interactive instruction were related to higher achievement score gains (about 0.05 above average).

Wenglinsky (2000)

Wenglinsky (2000) examined how teacher practices were associated with student achievement on the 1996 NAEP. Classroom practices were measured by survey during the administration of this national examination, particularly the use of small-group instruction and hands-on learning activities. The author also considered the contributions to student achievement of teacher education and experience as well as teacher professional development. He found that a teacher's major or minor in the subject taught correlated with higher student scores (0.09), as did the use of hands-on learning activities (0.25 for mathematics and 0.18 for science). Emphasis on higher-order thinking skills (0.13 for mathematics) was associated with increased student performance, but no significant differences were found for teachers' use of small-group instruction. Professional development that addressed working with special populations (0.21) and higher-order thinking skills (0.12) was related to higher scores in mathematics. Professional development in laboratory skills (0.13) was associated with higher scores in science. Lack of frequent point-in-time testing was related to lower scores in mathematics (–0.18), and regular assessment appeared to contribute to science scores (0.21).

Wenglinsky (2002)

Wenglinsky (2002) once again used NAEP survey data to examine how teachers' classroom practices, professional development, and qualifications (education level, mathematics subject area major or minor, and years of experience) relate to student achievement. He used multilevel structural equation modeling to distinguish between school- and student-level effects, to evaluate relationships among independent variables, and to model measurement error explicitly. Professional development in higher-order thinking skills and dealing with special populations was found to have significant effects. The school-level path model for classroom practices identified hands-on learning, solving unique problems, and avoiding reliance on authentic assessments as positively related to student achievement. The classroom practices investigated (20 variables) also were associated with higher student achievement.

Research on Teacher Effectiveness

Focus

The focus of this category of viewing teacher quality is on how linked teacher and student data are used to determine teacher effectiveness as measured by growth in student learning. Value-added measures are the most prominent of the methods used to assess teacher effectiveness in the studies summarized in this section.

Teacher effectiveness is becoming a topic of great interest among those interested in teacher quality. A number of policymakers and researchers have suggested that effectiveness, as measured by teachers' contribution to their students' learning, should be an important component of assessing teacher quality. Gordon, Kane, and Staiger (2006) wrote a discussion paper that uses their analysis of Los Angeles teacher-student linked data to suggest that teaching credentials matter little in terms of student achievement. The paper is interesting but is not included in the list below because it does not provide sufficient information about how the research design, methodology, and results meet the criteria for this research synthesis. The authors make a case for judging teachers on their effectiveness rather than on the basis of paper qualifications.

Findings

In general, these studies sought to demonstrate that differences in teacher effectiveness exist. In this goal, they generally were successful. However, taken as a whole, these studies were not able to arrive at convincing conclusions about which teacher qualifications, practices, or characteristics contributed to the differences in teacher effectiveness.

Research Studies

Aaronson, Barrow, and Sanders (2003)

Aaronson, Barrow, and Sanders (2003) conducted a study using Chicago public high school data with linked students and teachers. Using a value-added model and focusing on eighth- and ninth-grade standardized test scores for mathematics, the authors found that having an instructor who was rated two standard deviations higher than other teachers in quality (as determined by value-added scores) could add 25 percent to 45 percent of an average school year's growth to a student's mathematics score.

The authors also tried to correlate teachers' value-added scores with teacher characteristics for which they had data (age, experience, degree level, certification, and undergraduate major). They found that very little of the variance in teacher quality could be accounted for by these observable characteristics (except having an undergraduate major in mathematics or science). They concluded that teacher quality is largely attributable to characteristics not measured in this study. This finding suggests that variation in paper qualifications may matter little (with the exception perhaps of the undergraduate major, at least when mathematics is the subject being taught). Another implication from these findings is that what high-quality teachers do in their classrooms may be more important than their initial qualifications. Unfortunately, value-added

models tell nothing about *why* teachers vary in quality; nothing is known about their classroom activities that could help predict which teachers' students would gain the most.

As with all studies using value-added models, there may be issues with the interpretation of the findings. Some have argued that there is a problem with circularity in using value-added scores—which are largely based on student achievement scores—to then determine teacher contributions to student achievement gains (Kupermintz, 2003). Others have expressed concerns that what is being measured using such value-added models could more accurately be characterized as *classroom* effects, rather than *teacher* effects (Braun, 2004; National Association of State Boards of Education, 2005). The authors argue that it is not possible to separate the effects of other classroom-level contributors to student achievement using these models. For example, peer effects, availability of materials and books, school climate, and other effects could contribute to student learning at the classroom level, and these factors are largely outside the control of the teacher.

Noell (2006)

Noell (2006) used value-added scores for Louisiana students to examine the efficacy of teacher preparation programs. In the first phase of the research, value-added scores were calculated for students in Grades 4–9 in 66 of the 68 Louisiana public school districts, and then linked with teachers. Databases were constructed to allow separation of subject tests so that teacher effectiveness could be examined based on scores in specific subjects (English language arts, mathematics, science, and social studies). Not surprisingly, the single largest predictor of student achievement was the student's prior test score in the content area, followed by prior achievement in other subject areas. In the next phase of the study, teachers' preparation programs were identified and ranked according to estimates of effectiveness. Although the author found a relationship between teacher preparation programs and teacher effectiveness, large overlapping confidence intervals meant that the relationships could not be reliably determined with the data.

Nye, Konstantopoulos, and Hedges (2004)

Nye, Konstantopoulos, and Hedges (2004) wanted to determine the actual degree of teacher effects on student achievement. They defined teacher effects as the portion of student achievement that remains unaccounted for after controlling for student demographics, class size, and school fixed and random effects. To examine achievement gains, the authors also controlled for lagged test scores. The authors used data from the four-year Tennessee Project STAR (Student Teacher Achievement Ratio) experiment in which students and teachers were randomly assigned to classrooms with a range of teacher-pupil ratios. Their sample included 79 elementary schools in Tennessee. They found that between-classroom effects on achievement gains ranged from 0.123 (third grade) to 0.135 (second grade) for mathematics tests and from 0.066 (first grade) to 0.074 (third grade) for reading tests. All effects were significant. The between-classroom effects on achievement status were similar. The authors' examinations of teacher experience and education effects through hierarchical linear modeling, for the most part, were not significant or of small magnitude; some were even negative.

Rivkin, Hanushek, and Kain (2005)

Rivkin, Hanushek, and Kain (2005)[14] sought to sort out the impact of teachers (and schools) on achievement. Using matched panel data from Texas, the authors examined observable components (teacher education and experience) and unobservable components (residuals) and their relationship to student achievement gains on the Texas Assessment of Academic Skills in reading and mathematics. Focusing on Grades 3–7, the number of student scores ranged from 143,314 to 455,438 depending on the year and grade. The authors found that observable teacher characteristics have small but significant effects on student achievement gains but that most of teacher effectiveness is due to unobserved differences in instructional quality. They also determined that teacher effectiveness increased during the first year but leveled off after the third year.

Thum (2003)

Thum (2003) conducted research using linked archival data for elementary students and teachers in Arizona. He tested his production-function[15] model on 75 teachers and 1,276 students in Grades 3–6 in elementary schools in Arizona. He used student- and classroom-level covariates in the analysis, including sex, race, English proficiency, prior achievement, special education status, and grade level. He found that the mean growth for student test scores was positive and significant in all three grades. Using a teacher productivity profile (a function of targeted gains, degree of confidence, and model), he ascertained that only 17 of the 65 teachers who had 10 or more students in their classrooms achieved at least a 5 percent gain in student achievement in their classrooms at the 70 percent confidence level, and only 12 achieved that gain at the 80 percent confidence level.

Thum's findings suggest that while teachers are certainly contributing to student learning, it may be difficult to measure teachers' contributions with a high degree of certainty. Although many teachers had students who gained at least 5 percent, the confidence levels were too low to know whether such gains could be attributed to the teacher, to other sources, or merely to chance. For those who believe that teacher contributions to student learning are a measure of teacher quality, this question remains: How much confidence is enough for certainty that the gains are truly attributable to the teacher: 80 percent? 70 percent? less?

[14] This study also could be sorted with the Teacher Qualifications category because of the focus on teacher education and experience.

[15] An educational production function is a function in which a quantity of some educational input (such as years of teacher experience or per-pupil spending) yields a student output (such as test scores). There are many examples of production functions in the literature; good examples of the use of such functions include those by Hanushek, Kain, and Rivkin, (1998); Duncombe, Ruggiero, and Yinger (1996); and Monk (1994).

Summary and Recommendations

Challenges

Sensitivity of Measurement Tools

In some studies, factors that would logically be related to student achievement may appear to be only weakly related or not related at all. It might be a sample size issue because smaller sample sizes make it difficult to determine effects. Or it could be that the logic is wrong. But it also could be that the measurement tools and statistical analyses being conducted are not sensitive or precise enough to capture the effects. For example, statewide standardized student achievement tests are not ideal for measuring the effects of changes in instructional practice. Given that such tests occur once a year in most states and that teachers have the students in their classrooms for only six or seven months before the tests, subtle but important changes in practice may not show up as effects on achievement test scores. Similarly, increasing sophistication in database construction and the development of analytical approaches (such as hierarchical linear modeling) are rapidly changing the precision with which teacher effects are measured. However, it is likely that even better data systems and more precise statistical methods will be developed in the future.

Development of More Accurate Measurement Instruments

Another issue raised by evaluating these studies is that measurement instruments may not be appropriate for detecting subtle differences in teacher practices. For example, most of the scales used for teacher evaluation or for survey research are four-point Likert scales.[16] When a teacher is evaluated with such a scale, it is unlikely that he or she will score an average of 1 or 4. Instead, a teacher will probably score a few 1s, mostly 2s and 3s, and a few 4s. As a result, the average score will probably fall between 2.5 and 3.5. When the spread of the teacher's scores on this instrument is so constrained, it is very difficult to correlate the scores with student achievement and find meaningful, statistically significant effects. Thus, improving instruments to increase the range and precisions in scores from surveys and evaluations may produce more useful results.

Findings

Subject Matter and Grade-Level Differences in What Matters

The research highlighted in this synthesis clearly suggests that licensing for mathematics teaching and a degree in mathematics are positively correlated with mathematics achievement in all grades but particularly in secondary school. However, social studies, science, and other important school subjects have not been the focus of as much research as has mathematics. It remains to be seen whether subject-specific degrees and licensing in these other areas are essential for high levels of student learning.

[16] Likert scales indicate a level of agreement with a particular statement, usually on a four- or five-point scale from "strongly agree" to "strongly disagree." Problems with use of these scales include the tendency of respondents to (1) avoid the "extreme" answers and choose only the middle answers, and (2) be unwilling to answer in ways that might be considered "wrong" to others.

Why is mathematics apparently more sensitive to instruction in the classroom than reading? Nye, Konstantopoulos, and Hedges (2004) have theorized that "mathematics is mostly learned in school and thus may be more directly influenced by teachers.… Reading, on the other hand, is more likely to be learned (in part) outside of school" (p. 247). Thus, if students are exposed to mathematics concepts and given opportunities to explore and practice mathematics in only one place—the classroom—it is very important that the teachers be fully competent to guide their students' discovery. It may be less important for teachers in other subjects to have the kind of focused competence and course taking in their subjects. This finding suggests that tighter regulation of entry into mathematics teaching positions and more relaxed regulation of entry into teaching positions in other subjects might be appropriate.

In spite of the apparent importance of mathematics degrees and certification for student learning, there is an issue of supply and demand that must be resolved before moving to tighten requirements for mathematics teachers. Mathematics teachers are in short supply (National Commission on Mathematics and Science Teaching for the 21st Century, 2000; Office of Postsecondary Education, 2005; Urban Teacher Collaborative, 2000). The supply of mathematics teachers is unlikely to increase as long as there are (1) few salary incentives to become mathematics teachers, and (2) many salary incentives to go into other careers where mathematics skills are highly valued. If entry requirements into the teaching field are tightened for mathematics teachers, the supply of mathematics teachers may be reduced even more. Thus, there is an existing tension that must be resolved. One possibility is differential pay—paying properly trained and certified mathematics teachers salaries that are competitive with what they would earn if they took other career paths. This approach would be difficult to institute, however, given teacher organizations' lack of support for differential pay strategies.

Teacher Experience Matters, but Only in the First Few Years of Teaching

The finding that teachers reach their peak performance by increments within the first four or five years of teaching suggests that to continue efforts are needed to ensure that the most *inexperienced* teachers are not disproportionately assigned to schools where the challenges are greatest: schools with large percentages of low-income students, minority students, English language learners, and low-achieving students. As part of the NCLB highly qualified teacher requirements, states are under increasing pressure to ensure that highly qualified, *experienced* teachers are equitably distributed in schools. Few, if any, states have demonstrated that they have effective policies in place to ensure that beginning teachers are not disproportionately placed in hard-to-staff schools. But because of the pressure to demonstrate improvements in teacher distribution, states will be compelled to develop and implement a variety of strategies to address the problem. Evaluating the effectiveness of these strategies will be an important next step.

Another Consideration: Teaching Context

Another consideration is the context of the teaching. Should a teacher who is working in a challenging school with at-risk students be measured by the same yardstick as a teacher who is working in a high-achieving school in a middle-class suburb? Should the teaching context matter? Perhaps that is the wrong question. A better question might be as follows: "Within a given context—say, an at-risk urban school—what are the qualifications and characteristics

associated with teachers who are effective at producing student achievement?" By the same token, ask, "What are the practices that effective teachers in at-risk schools engage in that ensure high levels of student learning?"

Earlier in this synthesis, the point was made that there may be different definitions of teacher quality depending on the purpose at hand. Similarly, the set of inputs and processes that define teacher quality in one context may not be the same as those that define it in another. A highly successful, effective teacher in a suburban middle-class suburb may fail to be effective in an at-risk school in an urban setting, and vice versa. This situation does not mean that teachers should be judged by different standards according to their teaching context. Rather, it suggests the importance of learning more about what successful teachers do in every context.

Toward a New Definition of Teacher Quality

Given the significant advantages and disadvantages of the different mechanisms for determining teacher quality as reflected in these studies, it seems reasonable to suggest that a definition of teacher quality (and perhaps teacher certification) should encompass two components: (1) an initial set of qualifications tied to the subject matter and grade level being taught that must be met before a teacher is allowed to take charge of a classroom, and (2) some mechanisms for evaluating a teacher's effectiveness in producing student learning. With this combined definition, a two-stage process for assessing teacher quality may be needed: one based on paper qualifications and the other based on measures of teacher effectiveness that occur *after* the teacher has begun instructing students in the classroom. This assessment may involve some combination of expert or peer evaluation, teacher portfolios, and value-added scores.

Given the research analyzed through this framework, it seems apparent that defining teacher quality solely through paper qualifications is not sufficient for ascertaining teacher quality. Because the means are at hand to evaluate teachers' characteristics, practices, and effectiveness, reliance on paper qualifications as proxies for teacher quality is simply not sufficient for valid determinations of high- and low-quality teachers. This is not to say that paper qualifications—such as scores on a test of content knowledge—are useless. However, scores on tests cannot always predict which teachers will be most successful in the classroom. The challenge, therefore, is ensuring that licensure tests and other paper qualifications are in fact measuring what is most important: what the best teachers know and do that results in greater student learning in the classroom.

References

Aaronson, D., Barrow, L., & Sanders, W. (2003). *Teachers and student achievement in the Chicago public high schools* (Working Paper Series No. WP 02-28). Chicago: Federal Reserve Bank of Chicago.

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods, 10*(3), 317–328.

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., et al. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools* (No. R-2007-LAUSD). Santa Monica, CA: RAND Corporation.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman.

Associated Press. (2007, January 31). Houston doles out bonuses to teachers of core subjects. *Education Week, 26*(21), 8.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37–65.

Betts, J. R., Zau, A. C., & Rice, L. A. (2003). *Determinants of student achievement: New evidence from San Diego*. San Francisco: Public Policy Institute of California. Retrieved October 1, 2007, from http://www.ppic.org/content/pubs/report/R_803JBR.pdf

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal, 106*(1), 3–20.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2005). *How changes in entry requirements alter the teacher workforce and affect student achievement.* Albany, NY: Teacher Policy Research.

Braun, H. (2004). *Value-added modeling: What does due diligence require?* Princeton, NJ: Educational Testing Service.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models.* Princeton, NJ: Educational Testing Service.

California Department of Education. (1985). *Mathematics framework for California public schools: Kindergarten through grade twelve.* Sacramento, CA: Author.

Carr, M. (2006). *The determinants of student achievement in Ohio's public schools* (Policy Report). Columbus, OH: Buckeye Institute for Public Policy Solutions.

Cavalluzzo, L. C. (2004). *Is National Board Certification an effective signal of teacher quality?* (Report No. IPR 11204). Alexandria, VA: CNA Corporation.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). *Teacher-student matching and the assessment of teacher effectiveness* (NBER Working Paper No. 11936). Cambridge, MA: National Bureau of Economic Research.

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (CPRE Research Report No. RR-39). Philadelphia: Consortium for Policy Research in Education.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1).

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*(42). Retrieved October 1, 2007, from http://epaa.asu.edu/epaa/v13n42/v13n42.pdf

Darling-Hammond, L., & Youngs, P. (2002). Defining 'highly qualified teachers': What does 'scientifically-based research' tell us? *Educational Researcher, 31*(9), 13–25.

Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation.* Princeton, NJ: Mathematica Policy Research.

Dee, T. S. (2004). The race connection: Are teachers more effective with students who share their ethnicity? *Education Next, 2,* 52–59.

Duncombe, W., Ruggiero, J., & Yinger, J. (1996). Alternative approaches to measuring the cost of education. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 327–356). Washington, DC: The Brookings Institution.

Dwyer, C. A. (1994). *Development of the knowledge base for the Praxis III: Classroom performance assessments criteria*. Princeton, NJ: Educational Testing Service.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review, 48*(3), 547–561.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record, 107*(1), 186–213.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*(457), 465–498.

Ferguson, R. F., & Ladd, H. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd (Ed.), *Holding schools accountable: Performance based reform in education* (pp. 265–298). Washington, DC: The Brookings Institution.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7*(9). Retrieved October 1, 2007, from http://epaa.asu.edu/epaa/v7n9/

Frome, P., Lasater, B., & Cooney, S. (2005). *Well-qualified teachers and high-quality teaching: Are they the same?* (Research Brief). Atlanta, GA: Southern Regional Education Board. Retrieved October 1, 2007, from http://www.sreb.org/programs/hstw/publications/briefs/05V06_Research_Brief_high-quality_teaching.pdf

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4), 79–107.

Glass, G. V. (2005). Teacher characteristics. In A. Molnar (Ed.), *School reform proposals: The research evidence* (Chapter 8). Charlotte, NC: Information Age.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479–507.

Goe, L. (2002). Legislating equity: The distribution of emergency permit teachers in California. *Education Policy Analysis Archives, 10*(42), 1–50. Retrieved October 1, 2007, from http://epaa.asu.edu/epaa/v10n42/

Goldhaber, D., & Anthony, E. (2005). *Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching* (Working Paper). Seattle, WA: Center on Reinventing Public Education. Retrieved September 11, 2007, from http://www.crpe.org/workingpapers/pdf/NBPTSquality_report.pdf

Goldhaber, D. D., & Brewer, D. J. (1996, July). *Evaluating the effect of teacher degree level on educational performance*. Paper presented at the NCES State Data Conference.

Goldhaber, D. D., & Brewer, D. J. (1999). Teacher licensing and student achievement. In M. Kanstoroom & C. E. Finn, Jr. (Eds.), *Better teachers, better schools* (pp. 83–102). Washington, DC: The Thomas B. Fordham Foundation. Retrieved October 1, 2007, from http://www.edexcellence.net/doc/btrtchrs.pdf

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

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job: The Hamilton Project* (Discussion Paper 2006-01). Washington, DC: The Brookings Institution.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361–396.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373–388.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 61*, 280–288.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (Working Paper No. 11154). Cambridge, MA: National Bureau for Economic Research.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). *Teachers, schools, and academic achievement* (Working Paper No. 6691). Cambridge, MA: National Bureau of Economic Research.

Harbison, R. W., & Hanushek, E. A. (1992). *Educational performance of the poor: Lessons from rural northeast Brazil*. New York: Oxford University Press.

Harris, D. N., & Sass, T. R. (2007). *Teacher training, teacher quality and student achievement* (Working Paper No. 3). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved October 1, 2007, from http://www.caldercenter.org/PDF/1001059_Teacher_Training.pdf

Hattie, J. A. (1992). Towards a model of schooling: A synthesis of meta-analysis. *Australian Journal of Education, 36*, 5–13.

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

Heneman, H. G., Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (CPRE Policy Brief No. RB-45). Philadelphia: Consortium for Policy Research in Education. Retrieved October 1, 2007, from http://www.wcer.wisc.edu/cpre/publications/rb45.pdf

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 317–406.

Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(3), 207–219.

Jacob, B. A., & Lefgren, L. (2005). *Principals as agents: Subjective performance measurement in education* (Faculty Research Working Paper Series RWP05-040). Cambridge, MA: Harvard University.

Jesse, D., Davis, A., & Pokorny, N. (2004). High-achieving middle schools for Latino students in poverty. *Journal of Education for Students Placed at Risk, 9*(1), 23–45.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006, March). *What does certification tell us about teacher effectiveness? Evidence from New York City* (NBER Working Paper No. 12155). New York: National Bureau of Economic Research.

Kannapel, P. J., & Clements, S. K. (with Taylor, D., & Hibpshman, T.) (2005). *Inside the black box of high-performing high-poverty schools*. Lexington, KY: Prichard Committee for Academic Excellence. Retrieved October 1, 2007, from http://www.prichardcommittee.org/Ford%20Study/FordReportJE.pdf

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*(4), 54–78.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value-Added Assessment System. *Educational Evaluation and Policy Analysis, 25*(3), 287–298.

Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Education Policy Analysis Archives, 10*(37). Retrieved October 1, 2007, from http://epaa.asu.edu/epaa/v10n37/

Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science, 17*(3), 353–366.

Lockwood, J. R., Louis, T. A., & McCaffrey, D. F. (2002). Uncertainty in rank estimation: Implications for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics, 27*(3), 255–270.

Marcoulides, G. A., Heck, R. H., & Papanastasiou, C. (2005). Student perceptions of school culture and achievement: Testing the invariance of a model. *International Journal of Educational Management, 19*(2), 140–152.

Matsumura, L. C., Garnier, H., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment, 8*(3), 207–229.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., et al. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment* (CSE Technical Report No. 681). Los Angeles: Center for the Study of Evaluation.

McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education, 32*(5), 493–517.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.

McColsky, W., Stronge, J. H., Ward, T. J., Tucker, P. D., Howard, B., Lewis, K., et al. (2005). *Teacher effectiveness, student achievement, and National Board Certified teachers.* Arlington, VA: National Board for Professional Teaching Standards.

Mendro, R. L., Jordan, H. R., Gomez, E., Anderson, M. C., Bembry, K. L., & Schools, D. P. (1998, April). *An application of multiple linear regression in determining longitudinal teacher effectiveness*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33–53.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125–145.

Moore, W., & Esselman, M. (1992). *Teacher efficacy, power, school climate and achievement: A desegregating district's experience.* Paper presented at the American Educational Research Association, San Francisco.

Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review, 40*(2), 139–157.

National Association of State Boards of Education. (2005). *Evaluating value-added: Findings and recommendations from the NASBE Study Group on value-added assessments.* Alexandria, VA: Author.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation.* Washington, DC: U.S. Department of Education. Retrieved October 1, 2007, from http://www.ed.gov/inits/Math/glenn/report.pdf

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research. Retrieved October 1, 2007, from http://ccsr.uchicago.edu/publications/p0a02.pdf

Noell, G. H. (2006). *Value added assessment of teacher preparation* [Annual report]. Baton Rouge: Louisiana State University.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Office of Postsecondary Education. (2002). *Meeting the highly qualified teachers challenge* (The Secretary's Annual Report on Teacher Quality). Washington, DC: U.S. Department of Education. Retrieved October 1, 2007, from http://title2.ed.gov/ADATitleIIReport2002.pdf

Office of Postsecondary Education. (2005). *A highly qualified teacher in every classroom* (The Secretary's Fourth Annual Report on Teacher Quality). Washington, DC: U.S. Department of Education. Retrieved October 1, 2007, from http://title2.ed.gov/TitleIIReport05.pdf

Perkes, V. A. (1967). *Junior high school teacher preparation, teaching behaviors, and student achievement*. Unpublished doctoral dissertation, Stanford University.

Raymond, M., Fletcher, S. H., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas.* Stanford, CA: Center for Research on Education Outcomes. Retrieved October 1, 2007, from http://credo.stanford.edu/downloads/tfa.pdf

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes.* Washington, DC: Economic Policy Institute.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252.

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51–65.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education, 70*, 256–284.

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

Sanders, J. C. R. (1999). *The impact of teacher effects on student math competency achievement.* Unpublished dissertation, University of Tennessee–Knoxville.

Sanders, S. L., Skonie-Hardin, S. D., Phelps, W. H., & Minnis, T. L. (1994, November). *The effects of teacher educational attainment on student educational attainment in four regions of Virginia: Implications for administrators*. Paper presented at the annual meeting of the Mid-South Educational Research Association.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). *Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress*. Arlington, VA: National Board for Professional Teaching Standards. Retrieved October 1, 2007, from http://www.nbpts.org/UserFiles/File/SAS_final_NBPTS_report_D_-_Sanders.pdf

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247–256.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (No. R11-0435-02-001-97). Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Schacter, J., & Thum, Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review, 23*, 411–430.

Schlusmans, K. (1978). *What is an effective teacher?* Paper presented at the conference of the International Association for Educational Assessment, Baden, Austria.

Smith, J. B., Lee, V. E., & Newmann, F. M. (2001). *Instruction and achievement in Chicago elementary schools*. Chicago: Consortium on Chicago School Research. Retrieved October 1, 2007, from http://ccsr.uchicago.edu/publications/p0f01.pdf

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5*(1), 41–48.

Thum, Y. M. (2003). Measuring progress toward a goal estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods & Research, 32*(2), 153–207.

The Urban Teacher Collaborative. (2000). *The urban teacher challenge: Teacher demand and supply in the Great City schools*. Belmont, MA: Council of the Great City Schools. Retrieved October 1, 2007, from http://www.cgcs.org/pdfs/utc.pdf

Valentine, J., & Cooper, H. (2006). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes.* Washington, DC: What Works Clearinghouse.

Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National Board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46). Retrieved October 1, 2007, from http://epaa.asu.edu/epaa/v12n46/v12n46.pdf

Walberg, H. J., & Lai, J.-S. (1999). Meta-analytic effects for policy. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 419–453). New York: Academic Press.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). What helps students learn? *Educational Leadership, 51*(4), 74–79.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality* (Policy Information Center Report). Princeton, NJ: ETS. Retrieved October 1, 2007, from http://www.ets.org/Media/Research/pdf/PICTEAMAT.pdf

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*(12). Retrieved October 1, 2007, from http://epaa.asu.edu/epaa/v10n12/

Wiley, D. E., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis, 17*(3), 355–370.

Willett, J. B., Yamashita, J. J. M., & Anderson, R. D. (1983). A meta-analysis of instructional systems applied in science teaching. *Journal of Research in Science Teaching, 20*(5), 405–417.

Wilson, S. M., & Floden, R. (2003). Creating effective teachers: Concise answers for hard questions (Addendum to the report, *Teacher preparation research: Current knowledge, gaps, and recommendations.*) Washington, DC: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED476266). Retrieved October 1, 2007, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/0a/48.pdf

Wilson, S. M., Floden, R., & Ferrini-Munday, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations.* Seattle, WA: Center for the Study of Teaching and Policy. Retrieved October 1, 2007, from http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57–67.

Appendix A. Teacher Quality Variables Utilized

Teacher Effectiveness
Aaronson, Barrow, and Sanders (2003) Hanushek, Kain, O'Brien, and Rivkin (2005) Noell (2006) Nye, Konstantopoulos, and Hedges (2004) Rivkin, Hanushek, and Kain (2005) Rockoff (2004) Thum (2003)
Teacher Experience
Aaronson, Barrow, and Sanders (2003) Betts, Zau, and Rice (2003) Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005) Carr (2006) Cavalluzzo (2004) Clotfelter, Ladd, and Vigdor (2006) Darling-Hammond, Holtzman, Gatlin, and Heilig (2005) Decker, Mayer, and Glazerman (2004) Goldhaber and Brewer (1999) Hanushek, Kain, O'Brien, and Rivkin (2005) Harris and Sass (2007) Kane, Rockoff, and Staiger (2006) Leana and Pil (2006) Nye, Konstantopoulos, and Hedges (2004) Rockoff (2004) Rowan, Correnti, and Miller (2002) Wenglinsky (2002)
Teacher Education—*Certification*
Aaronson, Barrow, and Sanders (2003) Betts, Zau, and Rice (2003) Carr (2006) Darling-Hammond (2000) Darling-Hammond, Holtzman, Gatlin, and Heilig (2005) Decker, Mayer, and Glazerman (2004) Goe (2002) Goldhaber and Brewer (1999) Hanushek, Kain, O'Brien, and Rivkin (2005) Hill, Rowan, and Ball (2005) Kane, Rockoff, and Staiger (2006) Rowan, Correnti, and Miller (2002)

Teacher Education—*National Board Certification*
Cavalluzzo (2004)
Clotfelter, Ladd, and Vigdor (2006)
Goldhaber and Anthony (2005)
McColsky et al. (2005)
Sanders, Ashton, and Wright (2005)
Vandevoort, Amrein-Beardsley, and Berliner (2004)
Teacher Education—*Academic Major/Minor*
Aaronson, Barrow, and Sanders (2003)
Darling-Hammond (2000)
Harris and Sass (2007)
Monk (1994)
Wenglinsky (2000)
Wenglinsky (2002)
Teacher Education—*Advanced Degrees*
Aaronson, Barrow, and Sanders (2003)
Betts, Zau, and Rice (2003)
Carr (2006)
Cavalluzzo (2004)
Clotfelter, Ladd, and Vigdor (2006)
Darling-Hammond, Holtzman, Gatlin, and Heilig (2005)
Hanushek, Kain, O'Brien, and Rivkin (2005)
Harris and Sass (2007)
Nye, Konstantopoulos, and Hedges (2004)
Rowan, Correnti, and Miller (2002)
Wenglinsky (2000)
Wenglinsky (2002)
Teacher Education—*Preparation Experiences/Programs*
Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005)
Darling-Hammond, Holtzman, Gatlin, and Heilig (2005)
Decker, Mayer, and Glazerman (2004)
Frome, Lasater, and Cooney (2005)
Harbison and Hanushek (1992)
Harris and Sass (2007)
Hill, Rowan, and Ball (2005)
Kane, Rockoff, and Staiger (2006)
Noell (2006)
Teacher Education—*Undergraduate Institution Attended*
Aaronson, Barrow, and Sanders (2003)
Clotfelter, Ladd, and Vigdor (2006)

Teacher Education—*Teachers' Test Scores*
Cavalluzzo (2004)
Clotfelter, Ladd, and Vigdor (2006)
Harbison and Hanushek (1992)
Harris and Sass (2007)
Teacher Education—*Pedagogical Content Knowledge*
Betts, Zau, and Rice (2003)
Frome, Lasater, and Cooney (2005)
Harris and Sass (2007)
Hill, Rowan, and Ball (2005)
Monk (1994)
Rowan, Chiang, and Miller (1997)
Teacher Education—*Professional Development*
Harris and Sass (2007)
Wenglinsky (2000)
Teacher Evaluation Scores—*Standards-Based Ratings*
Borman and Kimball (2005)
Gallagher (2004)
Heneman, Milanowski, Kimball, and Odden (2006)
Holtzapple (2003)
Kimball, White, Milanowski, and Borman (2004)
Milanowski (2004)
Schachter and Thum (2004)
Teacher Evaluation Scores—*Principal Assessments*
Jacob and Lefgren (2005)
Instructional Practices
D. K. Cohen and Hill (1998)
Frome, Lasater, and Cooney (2005)
Kannapel and Clements (2005)
Marcoulides, Heck, and Papanastasiou (2005)
McCaffrey, Hamilton, Stecher, Klein, Bugliari, and Robyn (2001)
Rowan, Correnti, and Miller (2002)
Smith, Lee, and Newmann (2001)
Wenglinsky (2000)
Wenglinsky (2002)
Instructional Quality
Leana and Pil (2006)
Matsumura, Garnier, Pascal, and Valdés (2002)
Matsumura et al. (2006)
Newmann, Bryk, and Nagaoka (2001)
Smith, Lee, and Newmann (2001)

Teacher Attitudes—*Expectations for Students*
Frome, Lasater, and Cooney (2005) Rowan, Chiang, and Miller (1997)
Teacher Attitudes—*Teacher Collaboration*
Leana and Pil (2006) Rowan, Chiang, and Miller (1997)
Teacher Attitudes—*Teacher Efficacy*
Goddard, Hoy, and Hoy (2000) McColsky et al. (2005)
Teacher Race
Aaronson, Barrow, and Sanders (2003) Dee (2004) Ehrenberg, Goldhaber, and Brewer (1995) Hanushek, Kain, O'Brien, and Rivkin (2005)

Appendix B. Data Sources Used to Define Teacher Quality

Surveys—*Author-Developed Teacher Surveys*
Goddard, Hoy, and Hoy (2000)—teacher efficacy Hill, Rowan, and Ball (2005)—mathematical knowledge for teaching Leana and Pil (2006)—teacher social capital McCaffrey, Hamilton, Stecher, Klein, Bugliari, and Robyn (2001)—NCTM-aligned instructional practices McColsky et al. (2005)—efficacy Smith, Lee, and Newmann (2001)—instructional methods Vandevoort, Amrein-Beardsley, and Berliner (2004)—NBPTS status
Surveys—*NAEP Questionnaires*
Darling-Hammond (2000) Monk, 1994 Wenglinsky (2000) Wenglinsky (2002)
Surveys—*TIMSS Questionnaires*
Marcoulides, Heck, and Papanastasiou (2005)
Surveys—*National Education Longitudinal Study of 1988*
Ehrenberg, Goldhaber, and Brewer (1995) Goldhaber and Brewer (1999) Rowan, Chiang, and Miller (1997)
Surveys—*Prospects National Longitudinal Survey*
Rowan, Correnti, and Miller (2002)
Surveys—*California Basic Education Data System (CBEDS)*
Betts, Zau, and Rice (2003)
Surveys—*Brazilian EduRural Project*
Harbison and Hanushek (1992)
Surveys—*Student Surveys*
Frome, Lasater, and Cooney (2005)—teacher attitudes and instructional methods Marcoulides, Heck, and Papanastasiou (2005)—school culture
Surveys—*Parental Surveys*
Leana and Pil (2006)—satisfaction
Teacher Interviews
Jacob and Lefgren (2005) Kannapel and Clements (2005) Noell (2006) Nye, Konstantopoulos, and Hedges (2004) Rockoff (2004)

Instructional Artifacts
Matsumura, Garnier, Pascal, and Valdés (2002) Matsumura et al. (2006) McColsky et al. (2005) Newmann, Bryk, and Nagaoka (2001)
Observations—*Researcher*
Matsumura et al. (2006) McColsky et al. (2005) Schacter and Thum (2004)
Observations—*Administrator*
Borman and Kimball (2005) Gallagher (2004) Heneman, Milanowski, Kimball, and Odden (2006) Holtzapple (2003) Jacob and Lefgren (2005) Kimball, White, Milanowski, and Borman (2004) Milanowski (2004)
Observations—*Auditor*
Kannapel and Clements (2005)
Observations—*Based on Charlotte Danielson's Framework for Teaching*
Borman and Kimball (2005) Gallagher (2004) Heneman, Milanowski, Kimball, and Odden (2006) Holtzapple (2003) Kimball, White, Milanowski, and Borman (2004) Milanowski (2004)

Archival Data—*State or District Administrative Records*
Aaronson, Barrow, and Sanders (2003)
Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005)
Carr (2006)
Cavalluzzo (2004)
Clotfelter, Ladd, and Vigdor (2006)
D. K. Cohen and Hill (1998)
Decker, Mayer, and Glazerman (2004)
Darling-Hammond, Holtzman, Gatlin, and Heilig (2005)
Frome, Lasater, and Cooney (2005)
Goe (2002)
Goldhaber and Anthony (2005)
Harbison and Hanushek (1992)
Harris and Sass (2007)
Jacob and Lefgren (2005)
Kane, Rockoff, and Staiger (2006)
Kannapel and Clements (2005)
Leana and Pil (2006)
Noell (2006)
Rockoff (2004)
Sanders, Ashton, and Wright (2005)
Smith, Lee, and Newmann (2001)
Archival Data—*Tennessee Value-Added Assessment System*
Dee (2004)
Nye, Konstantopoulos, and Hedges (2004)
Archival Data—*Texas Schools Project*
Hanushek, Kain, O'Brien, and Rivkin (2005)
Rivkin, Hanushek, and Kain (2005)

Appendix C. Summary Table of Studies Examined[17]

Part 1. Research Studies Summarized in the Synthesis

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters[18]
Betts, Zau, and Rice (2003)	Teacher qualifications	Teacher credentials, education, experience, and subject-matter knowledge contributed to middle and high school student gains on the SAT-9, especially in mathematics, but actually *detracted* from elementary students' learning.
Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005)	Teacher qualifications	Teacher preparation, as reflected by traditional or alternative pathways into teaching, mattered to student gains on New York state achievement tests.
Carr (2006)	Teacher qualifications	"Highly qualified" designation of teachers made a small contribution to traditional public school students' Ohio state proficiency test gains, but not to charter school students' gains.
Cavalluzzo (2004)	Teacher qualifications	Teacher education, experience, and certifications, including National Board Certification, influenced student gains on Florida state proficiency tests in mathematics.
Clotfelter, Ladd, and Vigdor (2006)	Teacher qualifications	Teachers' experience, licensure test scores, and National Board Certification status mattered to North Carolina proficiency test scores in reading and mathematics, just mathematics, and just reading, respectively.
Darling-Hammond (2000)	Teacher qualifications	Teacher major and subject-area certification mattered to state-level NAEP mathematics and reading test scores.
Darling-Hammond, Holtzman, Gatlin, and Heilig (2005)	Teacher qualifications	Teacher certification influenced student achievement gains on TAAS, SAT-9, and Aprenda (a standardized Spanish-language test) in mathematics and reading.
Decker, Mayer, and Glazerman (2004)	Teacher qualifications	Teacher participation in Teach for America program influenced student achievement in mathematics.

[17] Many studies were examined in the course of identifying appropriate studies to include in this research synthesis but were not summarized in the narrative for a number of reasons, including length of time since the study was published, whether the methodology fit the criteria for inclusion (i.e., student achievement on standardized tests used as outcome variable), and whether the type of study (i.e., meta-analysis or descriptive study or summary that did not include data) was appropriate.

[18] Only those variables reported by the authors to have statistically significant associations are mentioned.

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters
Goe (2002)	Teacher qualifications	Teachers' certification status and experience mattered to school-level student achievement on California's Academic Performance Index (API).
Goldhaber and Anthony (2005)	Teacher qualifications	National Board Certification status contributed to students' state proficiency test scores.
Goldhaber and Brewer (1999)	Teacher qualifications	Teacher subject-matter certification (in mathematics) mattered to students gains on mathematics proficiency tests.
Hanushek, Kain, O'Brien, and Rivkin (2005)	Teacher qualifications	Teacher experience influenced student gains on TAAS.
Harbison and Hanushek, (1992)	Teacher qualifications	Teachers' education had a small, positive effect on Brazilian students' mathematics achievement; teacher scores on the same tests administered to their students mattered to Portuguese and mathematics test scores.
Harris and Sass (2007)	Teacher qualifications	Teachers' experience, content-oriented professional development, and pedagogical content knowledge predicted students' Florida Comprehensive Achievement Test (FCAT) scores, especially in middle school mathematics.
Hill, Rowan, and Ball (2005)	Teacher qualifications	Teachers' mathematical knowledge for teaching influenced student scores on the Comprehensive Test of Basic Skills (CTBS)/TerraNova mathematics test.
Kane, Rockoff, and Staiger (2006)	Teacher qualifications	While there were small differences in groups of teachers related to certification status (certified, uncertified, and alternatively certified), there were greater differences *within* groups, suggesting that variation among teachers' contributions to student achievement that is not accounted for by certification status. Teachers' contributions to student learning improves in the first few years of teaching.
McColsky et al. (2005)	Teacher qualifications	Teachers' National Board Certification did not predict students' achievement test gains.
Monk (1994)	Teacher qualifications	Teachers' subject-matter expertise, as reflected by academic course taking, contributed to students' NAEP mathematics and science scores but with diminishing returns after the fifth course.

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters
Rockoff (2004)	Teacher qualifications	Teacher experience influenced student CTBS and Metropolitan Achievement Test scores in mathematics and reading.
Rowan, Correnti, and Miller (2002)	Teacher qualifications	Teacher experience mattered to students' CTBS scores in reading and mathematics. Among practices also investigated, whole-class instruction and alignment of instruction with assessments mattered to students' learning.
Sanders, Ashton, and Wright (2005)	Teacher qualifications	Teachers' National Board Certification status did not reliably predict student achievement.
Vandevoort, Amrein-Beardsley, and Berliner (2004)	Teacher qualifications	National Board Certification status was associated with student gains on the SAT-9 tests.
Dee (2004)	Teacher characteristics	Student-teacher racial matching influenced SAT-9 scores in mathematics and reading.
Ehrenberg, Goldhaber, and Brewer (1995)	Teacher characteristics	Teachers' race, gender, and ethnicity did not contribute to students' scores on NELS assessments.
Goddard, Hoy, and Hoy (2000)	Teacher characteristics	Teachers' collective efficacy for teaching influenced students' Metropolitan Achievement Test scores.
Leana and Pil (2006)	Teacher characteristics	School internal social capital—defined as teachers' information sharing, trust, and shared vision—influenced student scores on state tests of mathematics and reading proficiency.
Borman and Kimball (2005)	Teacher practices	Teacher evaluations of instructional planning and instruction interactions (adapted from the Danielson [1996] *Framework for Teaching*) slightly influenced students' gains on Nevada proficiency tests and on CTBS/TerraNova tests.
D. K. Cohen and Hill (1998)	Teacher practices	Teachers' use of practices aligned to the 1985 *Mathematics Framework* contributed to student CLAS scores.
Frome, Lasater, and Cooney (2005)	Teacher practices	Teachers' instructional practices (such as group work on challenging assignments, oral and written reports on mathematics projects, and explaining solutions to the class) influenced Georgia proficiency test scores.

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters
Gallagher (2004)	Teacher practices	Teachers' literacy and composite evaluation scores (adapted from the Danielson [1996] *Framework for Teaching*) mattered to student achievement gains on the SAT-9.
Heneman, Milanowski, Kimball, and Odden (2006)	Teacher practices	Teachers' evaluation scores (adapted from the Danielson [1996] *Framework for Teaching*) influenced student achievement gains, especially when schools used trained and multiple observers.
Holtzapple (2003)	Teacher practices	Teachers' evaluation scores (adapted from the Danielson [1996] *Framework for Teaching*) mattered to student achievement on Ohio state proficiency tests.
Jacob and Lefgren (2005)	Teacher practices	Teachers' subjective assessments by principals contributed to student gains on core exams, especially in mathematics.
Kannapel and Clements (2005)	Teacher practices	Teachers' frequent assessments and feedback; use of student achievement data for staff development; instruction aligned to learning goals, assessments, and diverse learning styles; high expectations for student performance; and ongoing professional development differentiated between high- and low-performing high-poverty schools in Kentucky.
Kimball, White, Milanowski, and Borman (2004)	Teacher practices	Teachers' evaluation scores (adapted from the Danielson [1996] *Framework for Teaching*) contributed slightly to student gains on CTBS/ TerraNova and Nevada proficiency tests.
Marcoulides, Heck, and Papanastasiou (2005)	Teacher practices	Student reports that teachers assigned projects, had students discuss practical problems, assigned work relevant to students' daily lives, checked and discussed homework, and aligned the curriculum to assessments contributed to Greek students' achievement on TIMSS assessments.
Matsumura, Garnier, Pascal, and Valdés (2002)	Teacher practices	Teachers' use of high-quality assignments influenced student achievement on SAT-9 language arts tests.
Matsumura et al. (2006)	Teacher practices	Teachers' use of high-quality instruction and assignments mattered to achievement gains on some subscores of the SAT-10.
McCaffrey, Hamilton, Stecher, Klein, Bugliari, and Robyn (2001)	Teacher practices	Teachers' use of practices aligned to NCTM standards mattered to SAT-9 mathematics test scores for students in integrated/reform mathematics courses only.

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters
Milanowski (2004)	Teacher practices	Teachers' evaluation score (adapted from the Danielson [1996] *Framework for Teaching*) influenced student CTBS and state proficiency test scores in mathematics and reading.
Newmann, Bryk, and Nagaoka (2001)	Teacher practices	Teachers' use of intellectually demanding assignments mattered to Iowa Test of Basic Skills (ITBS) and state proficiency test scores.
Rowan, Chiang, and Miller (1997)	Teacher practices	Teachers' subject-matter knowledge, expectations for student outcomes, and placement in a collaborative school environment were associated with student achievement on NELS mathematics tests.
Schacter and Thum (2004)	Teacher practices	Teachers' performance ratings on 12 dimensions mattered to SAT-9 scores in mathematics, reading, and language arts.
Smith, Lee, and Newmann (2001)	Teacher practices	Teachers' use of interactive (rather than didactic) instruction contributed to student ITBS scores in mathematics and reading.
Wenglinsky (2000)	Teacher practices	Teachers' use of hands-on learning activities, emphasis of higher-order thinking skills, and professional development mattered to students' scores on NAEP assessments in mathematics and science.
Wenglinsky (2002)	Teacher practices	Classroom practices (especially hands-on learning, solving unique problems and *not* relying on authentic assessments) influenced NAEP mathematics test scores.
Aaronson, Barrow, and Sanders (2003)	Teacher effectiveness	Black box teacher "value-added" influenced students' ITBS to state proficiency test score gains in mathematics. A small amount of this variance could be attributed to the observed variable of undergraduate major.
Noell (2006)	Teacher effectiveness	Teachers' effectiveness varied according to which teacher preparation programs they attended, but relationships could not be determined with a high degree of certainty.
Nye, Konstantopoulos, and Hedges (2004)	Teacher effectiveness	Teacher effectiveness had small effects on student SAT-9 scores in mathematics and reading.
Rivkin, Hanushek, and Kain (2005)	Teacher effectiveness	Unobserved variables accounted for most of the difference in teacher effectiveness on students' gains on TASS.
Thum (2003)	Teacher effectiveness	Teacher effectiveness was difficult to measure with a high degree of certainty.

Part 2. Research Studies Not Summarized in the Synthesis

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters
Ballou, Sanders, and Wright (2004)	Not summarized (see footnote 17)	Black box teacher effectiveness mattered to students' gains on CTBS/TerraNova tests of reading, language arts, and mathematics. Lagged year test score was an appropriate proxy for student background variables.
Ferguson (1991)	Not discussed (see footnote 17)	Teachers' recertification exam scores influenced students' Texas Educational Assessment of Minimum Skills test scores and gains. Teachers' experience, up to five years, mattered to student dropout and SAT participation rates.
Ferguson and Ladd (1996)	Not discussed (see footnote 17)	Teachers' education mattered slightly to student test score gains, as did class size.
Fetler (1999)	Not discussed (see footnote 17)	Teachers' experience and certification status mattered to students' SAT-9 scores.
Goldhaber and Brewer (1996)	Not discussed (see footnote 17)	Teachers' mathematics and science subject-specific degrees influenced student test scores on NELS assessments of those subjects.
Good, Grouws, and Ebmeier (1983)	Not discussed (see footnote 17)	Teachers' use of active teaching process methods mattered differentially for the gains of students with low and high socioeconomic status (SES) on Science Research Associates tests of mathematics.
Greenwald, Hedges, and Laine (1996)	Not discussed (see footnote 17)	Teachers' ability and experience mattered to student achievement.
Hanushek (1971)	Not discussed (see footnote 17)	Teachers' verbal ability and recentness of education contributed to students' Stanford Achievement Test scores.
Hawk, Coble, and Swanson (1985)	Not discussed (see footnote 17)	Mathematics teachers' in-field certification mattered to students' gains on Stanford Achievement Tests of general mathematics and algebra.
Jesse, Davis, and Pokorny (2004)	Not discussed (see footnote 17)	A strong sense of guiding purpose, focus on achievement, supportive relationships among students and teachers, common goals, shared norms, consistent messages, and practices consistent with beliefs characterized nine Texas schools in which low-SES Hispanic students performed exceptionally well on TAAS.
Laczko-Kerr and Berliner (2002)	Not discussed (see footnote 17)	New teachers' certification status mattered to students' SAT-9 test scores.

Authors (Year)	Summary in Synthesis	How Teacher Quality Matters
Mendro, Jordan, Gomez, Anderson, Bembry, and Schools (1998)	Not discussed (see footnote 17)	Black box teacher effectiveness mattered to students' gains on the ITBS.
Mullens, Murnane, and Willett (1996)	Not discussed (see footnote 17)	Teachers' mathematics scores on their primary school exit exams and high school completion influenced students' gains on mathematics tests developed by the Belize Ministry of Education. Teachers' completion of pedagogic training did not significantly contribute to students' learning.
Perkes (1967)	Not discussed (see footnote 17)	Teachers' instructional methods, college grade point averages in science, and recentness of college-level study in science mattered to students' scores on tests of science knowledge and applications.
Raymond, Fletcher, and Luque (2001)	Not discussed (see footnote 17)	Teachers' Teach for America program participation contributed to students' gains on TAAS.
J. C. R. Sanders (1999)	Not discussed (see footnote 17)	Students' probability of passing the ninth-grade competency exam depended on the effectiveness of their teachers.
S. L. Sanders, Skonie-Hardin, Phelps, and Minnis (1994)	Not discussed (see footnote 17)	Teachers' educational degree level did *not* influence student dropout or postsecondary education enrollment rates.
W. L. Sanders and Rivers (1996)	Not discussed (see footnote 17)	Teacher effectiveness made both additive and cumulative contributions to students' gains on TCAP achievement tests.
Strauss and Sawyer (1986)	Not discussed (see footnote 17)	District average National Teacher Evaluation scores mattered to students' rates of failing state reading and mathematics competency exams.
Walberg and Lai (1999)	Not discussed (see footnote 17)	The effects of behavioral elements, teaching patterns, instructional systems, and teaching methods contributed to student outcomes according to this inclusive meta-analysis.
Wang, Haertel, and Walberg (1993)	Not discussed (see footnote 17)	Teachers' instructional quality and practices mattered to student achievement.
Wiley and Yoon (1995)	Not discussed (see footnote 17)	Teachers' implementation of instruction requiring higher level skills contributed to students' scores on CLAS tests.

Author (Year)	Summary in Synthesis	How Teacher Quality Matters
Willett, Yamashita, and Anderson (1983)	Not discussed (see footnote 17)	Teachers' use of innovative instructional systems (computer-simulated experiments, Bloom's Mastery Learning, and Keller's Personalized System of Instruction) was associated with students' science achievement.
Wright, Horn, and Sanders (1997)	Not discussed (see footnote 17)	Effective teachers contributed to the gains of students at all achievement levels—but especially to those of the lowest achieving group—on TCAP achievement tests.

Appendix D. Further Discussion of Effect Sizes

Effect size is commonly expressed with Cohen's *d*, where the difference of the means of the treated and control groups is standardized by dividing them by their pooled variance (J. Cohen & P. Cohen, 1983). This transformation results in an estimate of the magnitude of the effect in terms of standard deviations, allowing for meaningful comparisons of the size of different studies' results. In education research, typical effect sizes of interventions on student learning vary considerably, ranging from around 0.10 standard deviations to 0.70 standard deviations. In many cases, authors of research studies report the effect size; but when they do not, it sometimes can be approximated from the authors' reported test statistics (Grissom & Kim, 2005). Using effect sizes from various studies (either provided by the authors or calculated after the fact), a researcher developing a quantitative research synthesis can consider a key question: "Does the treatment help?" (Cooper & Hedges, 1994).

J. Cohen (1988) has pointed out that researchers often make the mistake of concluding that if a finding is statistically significant, it must be important. According to Cohen, what is really crucial is the *size of the effects*. As a rule of thumb, and accepted by many researchers today, he defined effect sizes as follows: small, $d = 0.2$; medium, $d = 0.5$; and large, $d = 0.8$. It is worth noting that Hattie (1992) conducted a meta-analysis of school interventions and discovered that simply spending a year in school has an effect size of approximately 0.40 on student learning (although that number may be somewhat inflated because it is difficult to take into account the length and intensity of the various interventions that were compared).

For an excellent discussion of some of the challenges related to calculating and reporting effect sizes in educational research, see Valentine and Cooper (2006). Progress is being made in understanding how best to calculate effect sizes, including an interesting alternative to Cohen's *d* (Algina, Keselman, & Penfield, 2005), and elaborations of some of the ways effect sizes are measured (Hancock, 2001). However, much more work remains in establishing the appropriate calculation of effect sizes before they will allow for true comparability across studies.