A POLICY INFORMATION CENTER REPORT

How Teaching Matters

Bringing the Classroom Back Into Discussions of Teacher Quality

ETS
Educational Testing Service

Milken Family Foundation

CONTENTS

Foreword *2*

Preface *3*

Acknowledgments *5*

Executive Summary *6*

Introduction *8*

Chapter One: Proposals to Improve Teacher Quality and What We Know About Their Effectiveness *10*

Chapter Two: A Portrait of America's Teachers and Their Classroom Practices *15*

Chapter Three: Linking Aspects of Teacher Quality to Student Test Scores *21*

Chapter Four: Conclusions *29*

Appendix: How the Study Was Conducted *32*

References *35*

This report was written by:

Harold Wenglinsky
Educational Testing Service

The views expressed in this report are those of the author and do not necessarily reflect the views of the officers and trustees of Educational Testing Service.

Additional copies of this report can be ordered for $10.50 (prepaid) from:

Policy Information Center
Mail Stop 04-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541-0001
(609) 734-5694
Internet – pic@ets.org

Copies can also be downloaded from:
www.ets.org/research/pic

October 2000

Foreword

Improving the quality of teaching in America's schools continues to be a central focus of educational reform. Increasing salaries, changing licensing requirements, and improving pre-service instruction are just some of the major initiatives targeted to bringing better prepared teachers into the classroom. Certainly, there is ample evidence, including the findings from this report, that well-prepared teachers produce more successful learners. Recruiting and preparing high quality teachers must remain a priority for policymakers.

Yet, Harold Wenglinsky, in this significant report, reminds us that attention only at the front end of the teacher pipeline is insufficient. Simply put, even if recruiting and preparation efforts were wildly successful, today's students do not have the luxury of waiting for a new generation of highly qualified teachers to staff our schools. For these students, it is imperative that their classroom teachers, today, are as effective as possible.

Wenglinsky's report is an optimistic one, giving us reason to believe that effective professional development does make a difference in student achievement. Of course, much of the professional development in our schools has been criticized as being superficial and irrelevant. What makes this report so important is that it begins to tease out the quality of professional development and its relationship to student learning. Professional development that attends to student learning of important skills and concepts appears to matter, and matter for all students in mathematics and science. Other kinds of professional development do not seem to matter. Wenglinsky has helped us move from the question "Should we support in-service professional development" to "What kinds of in-service professional development should be supported?"

As with many of his past reports, Wenglinsky capitalizes on the large-scale survey information from NAEP to give us insight into classroom practice and student learning. However, such data are limited in that they don't provide us a depth of understanding of the phenomena under investigation, in this case the detailed characteristics of professional development. But, these data certainly do give us reason to explore these issues with more intensive methodologies and smaller samples.

If we are to improve education, we must avoid the tendency to rely on simple generalizations and dichotomies. We need to attend to pre-service and in-service issues in improving teacher quality. We need to be discerning in the kinds of professional development that we support. And, we need to continue to do research like this that will illuminate our understanding of teacher practice and student learning.

On behalf of the Educational Testing Service, I would like to express my sincere appreciation to the Milken Family Foundation for supporting this work.

Drew Gitomer
Vice President - Research

Preface

The Milken Family Foundation has been studying school reform for over a decade. As we have considered specific reforms as diverse as early childhood education, smaller class sizes, parental involvement, and learning technology, one point became abundantly and unequivocally clear. Unless a child is taught by competent teachers, the impact of other education reforms will be diminished. Simply put, students learn more from "good" teachers than from "bad" teachers under virtually any set of circumstances. That is why the Foundation has developed the Teacher Advancement Program (TAP), the goal of which is to attract, motivate and retain the high quality educators we need to provide all children with the education they need and deserve.

The challenge to those recommending policy changes has been to define "good" teachers. In the simplest sense, good teachers are those who make students learn, and a better teacher will enable the same or similar students to learn more. Professor William Sanders, formerly of the University of Tennessee at Knoxville and now at the SAS Institute in North Carolina, along with a number of colleagues, has demonstrated the impact of good teachers. He sorts teachers according to the achievement of their students in a particular year. He then employs sophisticated statistical techniques to demonstrate that, year after year, students taught by the "effective" teachers (those whose initial students demonstrated high achievement) achieved more than students taught by other teachers.

Sanders has shown that the top 20% of teachers boosted the scores of low-achieving students over a one-year period by an average of 53 percentile points—39 percentile points higher than the 14 percentile point gain experienced by students assigned to the bottom 20% of teachers (Sanders & Rivers, 1998). Moreover, students who performed equally well in second grade showed a significant performance gap three years later depending on whether they had been assigned to the most effective or ineffective teachers. Fifth graders assigned to effective teachers in third, fourth, and fifth grades placed on average in the 83rd percentile. Those assigned to the least effective teachers over the same grades scored in the 29th percentile—a 54 point difference by the end of fifth grade (Sanders and Rivers, 1998:12). The Sanders work and similar research around the country is regularly cited to demonstrate the powerful impact of "good" teachers.

But Sanders leaves us with a "black box" in regard to the meaning of a good teacher. His work tells us that good teachers get good results for students. But are good teachers prepared differently than others? Are they inherently smarter? Do good teachers teach differently or use different teaching methods in their classes? Do teachers get higher student achievement when they are rewarded for doing so—financially or otherwise?

As Dr. Harold Wenglinsky discusses in this monograph, researchers have tried to answer most of these questions for decades. Their findings have been inconclusive or contradictory, mostly because of deficiencies in their data or statistical methods. Perhaps the most enduring findings in this regard have been that student achievement is higher, *ceteris paribus*, when teachers have high verbal skills (Ehrenberg & Brewer, 1995; Ballou & Podgursky, 1997) and when they have strong knowledge of the discipline they teach (usually indicated by a major or minor in that field) (Goldhaber & Brewer, 1999; Fetler, 1999; Monk, 1994).

What we have not known, until now, is whether good and effective teachers do things in their classrooms differently than less effective teachers. Some earlier small scale qualitative or quantitative studies have been suggestive, but Dr. Wenglinsky's work presented here is the first attempt (using a national database and sophisticated analytical techniques) to answer the question of whether effective teachers do things differently in their classrooms.

By controlling for some teacher inputs, student characteristics, and school measures, Dr. Wenglinsky demonstrates that certain types of professional development prepare teachers to use specific techniques in their classrooms that result in higher student achievement.

Dr. Wenglinsky finds that only one teacher input, majoring or minoring in the subject taught, is associated with improved academic performance. This is consistent with earlier findings. Although teachers generally give their professional development activities mixed reviews, this study suggests that certain components of such training are useful. In particular, professional development in working with special populations, in higher-order thinking skills for math, and in laboratory skills for science are associated with better student performance. These findings must be kept in mind as the professional development plans for TAP are designed.

The findings regarding professional development are supported by what works in the classroom. Students whose teachers emphasize higher-order thinking skills (math) and hands-on learning activities (e.g., lab work in science) outperform their peers significantly. Also, students who frequently take tests outperform those frequently using on-going forms of assessment such as portfolios. Thus Dr. Wenglinsky's study confirms the Foundation's view that what goes on in classrooms matters—effective teachers do things differently.

There are four school-related principles that guide TAP's efforts to attract, motivate, and retain high quality teachers to K-12 education.

With *multiple career paths*, TAP offers all teachers the opportunity to advance in the profession without having to leave the classroom. Teachers are able to move along a continuum ranging from inductee to master teacher where increased responsibilities, qualifications, professional development, and performance-based accountability requirements are commensurate with compensation. Multiple career paths provide for expanded roles for talented teachers as leaders, decision-makers, and mentors at the school site, and opportunities to work in the broader community.

Market-driven compensation replaces lock-step salary structures and provides flexibility to establish salaries. Pay differentiation is based on demand (more to those in hard-to-staff fields and schools), demonstrated teacher knowledge and skills, actual teacher performance, increased responsibilities, and student performance. This system provides increased pay for those who do more work and are judged to be the best.

Performance-based accountability is rigorous, tied to compensation, and includes differentiated requirements based on the teacher's position. Teachers are assessed against high standards that measure their performance in content knowledge, planning, instruction, assessment, and in producing student learning gains. Hiring, advancement, and pay increases are based on performance reviews conducted by the principal and peer experts from both within and outside of the school. The ultimate goal is for teachers to sign three-year renewable contracts. In the short term, tenure reviews will be more thorough, and tenure will be awarded after a longer period of time in the classroom. Initial and continuing certification will become primarily performance based.

Ongoing, applied professional growth requires a school-wide commitment and includes all teachers. Outcomes are tied to state teaching and learning standards, school improvement efforts, and a data-driven analysis of student learning outcomes. Activities occur at the school site during frequent professional growth blocks led by the principal and master teachers and guided by mentor teachers; they are designed to encourage more collaboration among professional staff. A mandated salaried, mentored induction year gives new teachers classroom responsibility with intensive support. Ongoing, applied professional growth ensures adequate time for teachers to meet, reflect, learn, and grow professionally.

In addition, TAP has a fifth principle, *expanding the supply of high quality teachers*. This is achieved by making the initial academic degree and teaching certification attainable in four years; providing alternative certification to give beginning teachers as well as mid-career professionals the ability to enter teaching as adjuncts or full-time through assessments and classroom demonstration; and allowing outstanding retired teachers to continue working on a part-time basis as faculty fellows. Expanded teacher job mobility—and, therefore, increased competition and opportunity for teachers—is achieved through multi-state credentialing, private pension plans that make benefits more portable, and the opportunity for all teachers to become nationally certified at the beginning, middle, and advanced levels of professional practice.

Dr. Wenglinsky's study helps us implement a number of these principles. It suggests particular types of professional development that can help teachers enhance student achievement. It also informs TAP's assessment model by suggesting specific classroom activities that should be evaluated. And it gives weight to our performance-based compensation system in that it shows that we can reward particular behaviors identified as being associated with greater student achievement.

Dr. Wenglisky's study is the first of its kind, and it may not be the final word. But the Milken Family Foundation is pleased to have sponsored it because it should be a catalyst for further research and discussion about what makes good, or even great teachers.

Lewis C. Solmon
Senior Vice President and
Senior Scholar
Milken Family Foundation

Acknowledgments

This report and the study on which it is based received assistance from many quarters. The study was funded by a generous grant from the Milken, Family Foundation. Lewis C. Solmon, Senior Vice President and Senior Scholar at Milken, provided invaluable intellectual guidance through all stages of the project, from its initial conception to the completion of the report. The data for the statistical analyses were prepared by Alfred Rogers. The report was reviewed by Paul Barton, Robin Henke, and Jacqueline Jones. Richard Coley provided the graphics. Carla Cooper provided desktop publishing services. Marie Collins was the editor. Marita Gray designed the cover. Kathleen Benischeck coordinated production. This report does not, however, necessarily represent the views of those who contributed to its production. Any errors of fact or interpretation are the author's responsibility.

Executive Summary

Improving teacher quality has become the subject of numerous policy proposals at the federal and state levels. In the wake of efforts to raise academic standards across the country, policymakers have expressed concern that many teachers are not prepared to help students meet the new standards. Proposals to remedy this situation have included increasing teacher salaries to attract better qualified teachers, requiring more education, such as a master's degree, or requiring a major in the subject a prospective teacher plans to teach.

These proposals tend to focus on improving nonclassroom aspects of teacher quality, often neglecting the nature of the teaching and learning that occurs within the classroom. Most proposals either make teaching more lucrative, in the hope of recruiting better teachers, or raise the bar for those going into teaching through more stringent licensing requirements. The assumption these proposals make is that such efforts will improve teacher quality, resulting in improved student academic performance. Unfortunately, the empirical evidence on this contention is extremely mixed. Research has not consistently demonstrated a link between teacher inputs, such as salaries and education levels, and student outcomes, such as scores on standardized tests.

The study presented in this report explores another route to improving teacher quality—namely, improving teachers' classroom practices. Teachers employ a variety of strategies—from problem sets to small-group activities—to encourage students to learn. Some of these strategies are presumably more effective than others. Yet little is known about which strategies are most effective, because large-scale studies relating classroom practices to student test scores or other outcomes have almost never been conducted. Without a stronger research base for identifying effective classroom practices, policymakers will find it difficult to develop policies that encourage these practices.

The current study represents a first step toward linking classroom practices to student academic performance. It does so by analyzing data from the National Assessment of Educational Progress (NAEP). Known as "the Nation's Report Card," NAEP is administered to students across the nation every year or two in a variety of subjects. In addition to standardized tests consistent with high academic standards, the NAEP database includes questionnaires sent to students, their teachers, and their principals. From these data it is possible to relate various aspects of teacher quality to student test scores, while taking into account other potential influences on these scores, such as class size and student social background. For this study, data are examined for two national samples of students:

- 7,146 eighth graders who took the NAEP mathematics assessment in 1996
- 7,776 eighth graders who took the NAEP science assessment in 1996

Three types of teacher quality are measured:

- teacher inputs, such as teacher education levels and years of experience
- classroom practices, such as the use of small-group instruction or hands-on learning
- professional development, meaning training to support certain classroom practices

The study begins by describing the national picture for teacher inputs, professional development, and classroom practices. Some key findings are:

- Among teacher inputs, in both math and science, one out of three teachers has at least a master's degree; three out of four majored or minored in the subject they are teaching; and six out of 10 have at least 10 years of teaching experience.
- In both math and science, approximately half of all teachers have received more than two days of professional development in the last year.
- In both math and science, the most common topic for professional development is cooperative learning.

- In math, the least common topics for professional development are those dealing with special student populations, such as those with limited English proficiency or special needs.

- In science, the least common topics for professional development are those dealing with special student populations and those dealing with laboratory skills.

- In math, teachers are more likely to prepare students to answer routine problems than to answer problems involving new or unique situations.

- Less than one out of four math teachers engages in hands-on learning activities with their students, whereas two out of three science teachers report engaging in such activities.

- While nearly all math and science teachers report testing students at least once a month, these tests are more likely to involve extended written answers than multiple-choice responses.

The study then links these classroom practices, professional development experiences, and teacher inputs to student academic performance. It finds:

- Of the teacher inputs, only majoring or minoring in the relevant subject is associated with improved student academic performance. Students whose teachers majored or minored in the subject they are teaching outperform their peers by about 40% of a grade level in both math and science.

- In math, students whose teachers have received professional development in working with special populations outperform their peers by more than a full grade level, and students whose teachers have received professional development in higher-order thinking skills outperform their peers by 40% of a grade level.

- In science, students whose teachers have received professional development in laboratory skills outperform their peers by more than 40% of a grade level.

- In math, students whose teachers emphasize higher-order thinking skills outperform their peers by about 40% of a grade level.

- Students whose teachers conduct hands-on learning activities outperform their peers by more than 70% of a grade level in math and 40% of a grade level in science.

- Students who frequently take point-in-time tests outperform those frequently using on-going forms of assessment, such as portfolios, by 46% of a grade level in math and 92% of a grade level in science.

The last of these findings should not be taken to mean that portfolio assessments should be entirely supplanted by multiple-choice tests. For one, portfolios may have important functions other than monitoring the progress of individual students, such as providing professional development for teachers and information on the progress of entire classes. For another, the tests that benefit students are not necessarily of a multiple-choice format. Rather, both tests involving multiple-choice responses and those involving extended written responses proved effective.

Overall, these findings suggest that policymakers are correct in emphasizing the importance of improving teacher quality as a mechanism for improving student academic performance. However, these findings indicate that greater attention needs to be paid to improving classroom aspects of teacher quality. In particular, teachers should be encouraged to convey higher-order thinking skills, conduct hands-on learning activities, and rely primarily upon tests to monitor student progress. Policymakers can encourage these practices by providing rich and sustained professional development that is supportive of these practices, and provided that teachers have access to such professional development, perhaps by rewarding them either financially or through advanced forms of certification for engaging in these practices.

Introduction

Teacher quality has come to the fore in discussions of educational reform. During the 1980s and much of the 1990s, policymakers focused on setting high academic standards for all students in elementary and secondary schools. The focus started changing with the release of the Hunt Commission report, *What Matters Most: Teaching for America's Future,* in 1996 (National Commission on Teaching and America's Future, 1996). The report contended that, for students to meet high academic standards, the quality of the teaching force needed to be improved. A flurry of proposals, pronouncements, and legislation on improving teacher quality followed. While the approaches varied somewhat, they had in common the belief that improving teacher quality was the next step that needed to be taken to improve education.

The approaches had another thing in common: They tended to stress aspects of teaching that occurred outside of the classroom, known as teacher inputs, as the instrument for improving student performance. The nonclassroom factors manipulated by these proposals included financial aspects of teaching (such as salaries, tax abatements, or bonuses for high performance) and the qualifications of teachers (such as requiring teachers to pass a licensure examination or encouraging teachers to obtain a master's degree). In manipulating teacher inputs, policymakers assumed that there was an empirical link between these inputs and the outputs of education, in particular, student test scores and other measures of academic performance. Unfortunately, this assumption is not necessarily tenable; research on the link between teacher inputs and student outputs has been inconclusive. In addition, the focus on nonclassroom factors neglects the important role that the practices actually occurring in the classroom may play in student learning. In addition to being the mechanism through which improvements in teacher inputs might translate into higher student achievement, classroom practice may influence student achievement irrespective of the finances and qualifications of the teachers entering the classroom, and may even influence student achievement more strongly than teacher inputs.

The purpose of the current study is to explore the possible influence of classroom practices on student achievement. The study uses data from the National Assessment of Educational Progress (NAEP), a national sample of students and their schools. From this database, it is possible to measure student scores on assessments of mathematics and science, a comprehensive set of classroom practices, the training teachers receive that is specifically tailored to these practices (known as professional development), and various teacher inputs. In addition, information is available from the NAEP database on other characteristics of schools and students that may potentially influence test scores, such as class size. By analyzing these data, the study can measure the impact of classroom practices, professional development, and teacher inputs relative to one another. It can also place classroom practices in the process through which other aspects of teacher quality affect student learning; a flow chart can be developed that measures the extent to which teacher inputs and professional development may influence classroom practices that may in turn influence student academic performance. The NAEP data thus make it possible to draw together more information about the relationship between classroom practices and student academic performance than has ever existed before for a national sample of students.

The study finds that while teacher inputs, professional development, and classroom practices all influence student achievement, the greatest role is played by classroom practices, followed by professional development that is specifically tailored to those classroom practices most conducive to the high academic performance of students. In particular, when teachers make use of hands-on activities to illustrate concepts in mathematics and science, students perform better on assessments in these subjects. Also, when teachers focus on conveying higher-order thinking skills, particularly those involving the development of strategies to solve different types of problems, students perform better on mathematics assessments (though no better on science assessments). The study also finds that the methods teachers use to assess student progress have a similar impact on achievement in mathematics and science. Tests given at a particular point in time are associated with higher student performance than on-going techniques, such as portfolio or project-based assessment. Professional development activities in hands-on learning and higher-order thinking skills are also associated with improved student performance. Of the teacher inputs studied, the only one that makes a difference is the teacher's major or minor in college. Students do better when their teachers major or minor in the subject areas they teach. Overall, professional development is three times as important as teacher inputs in mathematics and one and a half times as important in science. Classroom practices are five times as important as teacher inputs in mathematics and four times as important in science.

Before discussing these findings in more detail, however, it is necessary to provide some background. Chapter One presents some of the major policies to improve teacher quality that have been proposed by federal and state officials. It then summarizes what is known from prior research about teacher inputs, professional development, and classroom practices, as well as their impact on student academic performance. Chapter Two presents descriptive information from NAEP on the prevalence of various teacher inputs, professional development, and classroom practices in U.S. schools. Chapter Three then uses statistical techniques to relate these aspects of teacher quality to student performance in mathematics and science. Chapter Four teases out some implications of this study for policies on teacher quality, and suggest directions for further research.

Chapter One: Proposals to Improve Teacher Quality and What We Know About Their Effectiveness

Proposals and some actual legislation to improve teacher quality have come fast and furious over the last few years. At the federal level, President Clinton has given visibility to the issue in State of the Union addresses. In February 1999, Secretary of Education Richard Riley unveiled the Clinton Administration's program for improving teacher quality. It called for states to institute three levels of licensure for teachers. Under this system, an initial license to teach would be granted to prospective teachers who could pass written examinations in the subject they planned to teach, knowledge of pedagogy, and basic skills. Within a few years, teachers would be expected to obtain professional licenses. These licenses would be granted based upon observations of classroom performance by fellow teachers. Still later in their careers, teachers could be evaluated for advanced licenses along the lines of the National Board for Professional Teaching Standards (NBPTS) system, which involves a protocol for collecting and scoring classroom observations, portfolios of student work, and other kinds of information about teachers.

In addition to proposing the three-step licensure system, Riley proposed a national job bank through which school districts could recruit teachers, a national conference on teacher quality, and a national commission on mathematics and science teaching (Robelen, 1999).

Teacher quality was also the central topic of discussion at the third educational summit in October 1999. Prior summits had focused on national academic standards and how to measure student progress toward those standards. The third summit identified improving teacher quality as a crucial mechanism for enabling students to meet high standards, and pointed to some possible tools for improving teacher quality. These tools included providing bonuses to teachers in exchange for high academic performance among their students; offering teachers professional development closely linked to high academic standards; and permitting pathways to licensure other than through the traditional undergraduate-level education programs based in colleges and universities (Olson & Hoff, 1999).

This year, the candidates for the 2000 presidential election have offered their own proposals for improving teacher quality. Al Gore has called for 60,000 scholarships for students who commit to teaching in schools in particular need of teachers for at least four years; bonuses and training for 15,000 mid-career professionals who want to become teachers; and "Higher Standards, Higher Pay" grants to help urban and rural districts attract and retain high quality teachers by offering higher salaries and other benefits. George W. Bush has proposed $2.9 billion in new spending to recruit, hire, and train teachers; tax incentives to help teachers pay for school supplies and other school-related out-of-pocket expenses; and increased funding for Troops to Teachers, a program to move members of the Armed Services into teaching.

Proposals have been proliferating among state officials as well. Governor Gray Davis of California has signed legislation providing tax benefits to teachers as well as raising their salaries. San Francisco plans to offer its teachers subsidized housing. In Massachusetts, mathematics teachers in schools where more than 30% of the students fail the state mathematics assessment will be required to take a mathematics proficiency test to be recertified as teachers. In New York, Governor George Pataki is offering financial rewards to students in the state college and university system who agree to teach in districts or subjects in which teachers are in short supply. Washington State has recently become the 40th state to require new teachers to take tests in their subject area and in knowledge of pedagogy; the tests are to be overseen by a new professional standards board. And in Kentucky, legislation was defeated that also would have created a state professional standards board, as well as required middle school teachers to demonstrate proficiency in their subject area. Instead, the legislation which passed expanded summer school training of middle school teachers and established a new system for evaluating teacher training programs (e.g., Bradley, 2000; Hoff, 2000).

Despite all of the attention that improving teacher quality has garnered, the policies that have been proposed generally suffer from two problems. The proposals each involve manipulating a particular teacher input in the hope that changes in the input will automatically lead to improvements in an output—namely student academic performance. One problem with this approach is that there is little empirical evidence of a link between improvements in teacher inputs and improvements in student outputs. A second problem is that, in emphasizing aspects of teacher quality that occur outside of the classroom, these proposals often ignore practices in the classroom. Yet changing the nature of teaching and learning in the classroom may be the most direct way to improve student outcomes.

Research on the links between teacher inputs and student outputs began nearly 35 years ago with the publication of the Equality of Educational Opportunity Study of 1966, also known as the Coleman Report (Coleman et al., 1966). This study collected information on a series of teacher inputs, such as teacher education levels and years of teaching experience, and a series of student outputs, such as scores on standardized tests, for a national sample of students. The study then related the inputs to the outputs, and found that, in most instances, the link between the two was weak. In the wake of the Coleman Report, about 400 studies have been conducted relating teacher inputs to student outputs. Their results have been extremely mixed. Of those studies linking the number of years of teaching experience to student outcomes, 30% showed a beneficial effect. Of those studies linking the salaries of teachers to student outcomes, 20% showed a beneficial effect. And of those studies linking the education levels of teachers (master's degrees as opposed to bachelor's degrees), 10% showed a beneficial effect. Thus, while a substantial number of studies did find evidence of a link between a given input and student outputs, most did not. Not only do the studies disagree on the efficacy of teacher inputs, but studies that attempt to tabulate and interpret the results of these studies come to different conclusions. Some suggest that these studies provide sufficient evidence that teacher inputs can be important; others contend that stronger evidence would be needed to make that claim (Hanushek, 1997; Greenwald, Hedges & Laine, 1996).

One exception to the mixed findings of input-output research has been studies of teacher skills as measured by standardized tests. The original Coleman Report found a strong link between teacher scores on a vocabulary test and student academic performance. More recently, Ferguson (1991) found that when teachers perform well on basic skills tests, their students do better as well. And Wright, Horn, and Sanders (1997) found that teacher test scores are strongly related to improvements in student test scores over the course of a year (referred to as the "value added" by the teacher to the student).

The mixed findings of the input-output studies can be accounted for in two ways. It may be that the link between teacher inputs and student test scores is simply weak and mixed; it would thus be expected that some studies might pick up a little bit of a link while others do not. Or, the mixed findings may reflect problems in study design. Other than the Coleman Report, most of these studies were small in scale, consisting of a single state or even a single school district. Thus, the mixed findings may reflect the fact that certain inputs may be important in certain places and not others. Another design problem is that the studies often lack rich measures of the output. They may measure student outcomes based upon a minimum skills test, for instance, rather than through a standards-based test that requires students to demonstrate mastery of critical thinking skills. These studies also often lack measures of inputs, other than teacher inputs, that may have an independent effect on student test scores, such as the social background of students.

And, many of these studies fail to take into account what is referred to as the "multilevel" nature of input-output studies. That is, such studies relate an input that is at one level of aggregation—the classroom or its teacher—to an output that is at a lower level of aggregation—each student in each classroom. To the extent that each of these problems occurs in the study design, results may vary.

Thus, current policy proposals seek to improve student academic performance by tapping aspects of teacher quality, namely aspects outside of the classroom, when there is little research support for their efficacy. In addition, these proposals ignore other aspects of teacher quality, namely aspects within the classroom which may be more promising.

For the most part, researchers of the links between teacher quality and student performance tend to ignore classroom practices as much as policymakers do. To date, there is little quantitative research that relates a comprehensive list of classroom practices to student achievement, let alone also taking into account professional development and teacher inputs. Nonetheless, prior research does provide some useful information on the issue. There have been a host of qualitative studies of classroom practices that, largely through the technique of classroom observation, identify different sets of practices common to teachers and suggest the effectiveness of the different sets. There have also been quantitative studies of classroom practices which—while they do not relate these practices to student outcomes—do provide useful information on their prevalence. Finally, there has been one quantitative study that related a set of classroom practices to student performance in mathematics and science. While the set of practices was rather limited, and other aspects of teacher quality, such as professional development, were not taken into account, this study does provide some suggestive information upon which the current study can build.

Qualitative research on classroom practice has put forward certain propositions regarding which practices are most effective—propositions that can be tested through quantitative research. First, it asserts the importance of teaching higher-order thinking skills. Very often, teaching involves conveying information to students. Students are expected to retain this information, and learn to do so either through memorization or by solving problems that are similar to one another. Such techniques for knowledge acquisition are often referred to as drill and practice, learning by rote, or lower-order thinking skills. In other cases, teaching involves not so much conveying information as conveying understanding. Students learn concepts and then attempt to apply them to various problems, or they solve problems and then learn the concepts that underlie the solutions. This movement between the abstract and the concrete is referred to as "higher-order thinking skills," or alternately as "critical thinking skills." These skills tend to be conveyed in one of two ways: through applying concepts to problems (applications) or by providing examples or concrete versions of the concept (simulations). In either case, students learn to understand the concept by putting it in another context. In the case of an application, this might mean solving a unique problem with which the student is unfamiliar. In the case of a simulation, this might mean examining a physical representation of a theorem from geometry or engaging in a laboratory exercise that exemplifies a law from chemistry. While both lower-order and higher-order thinking skills undoubtedly have a role to play in any classroom, much of the qualitative research asserts that the students of teachers who can convey higher-order thinking skills as well as lower-order thinking skills outperform students whose teachers are only capable of conveying lower-order thinking skills (McLaughlin & Talbert, 1993).

Second, the qualitative research suggests a few teaching methods that are conducive to the teaching of higher-order thinking skills. Individualization of instruction is important. Students bring different bodies of knowledge with them into the classroom. Rather than treat all students as blank slates, teachers should instruct each student by drawing upon the knowledge and experience that particular student already has.

Collaborative learning is also important. Teachers should allow students to work together in groups, and teachers should work collaboratively with students at an individual, small group, and whole class level to bring out key concepts and problem-solving techniques, rather than have all ideas originate from the teacher. Finally, student progress should be assessed in an on-going fashion rather than at discrete points of time. Assessment should be no different than the student's ordinary activities for learning. It is important not to disrupt the learning environment with artificial activities, such as tests.

The prevalence of many of these practices as well as the prevalence of many nonclassroom aspects of teacher quality has been measured in national surveys of America's teachers. A recent survey of teachers undertaken by the U.S. Department of Education provides a snapshot of the distribution of teacher inputs across the nation (National Center for Education Statistics, 1999a). One teacher input frequently discussed is the education level of teachers; many reform proposals call for greater numbers of teachers or prospective teachers to obtain master's degrees. The Department of Education study indicates that, nationwide, 45% of teachers hold master's degrees. Teachers with master's degrees are disproportionately located in the Northeast (60% in this region have master's degrees), teach high school (55%), have 20 or more years of teaching experience (62%), and have predominantly affluent students (57% teach students in schools where less than one in six students qualifies for a free or reduced price lunch). Another frequently discussed input is the undergraduate major of the teacher. The Department of Education study indicates that 38% of teachers major in an academic field, 24% in a specific subject area in education (such as mathematics education), and the remainder in education not tied to a specific subject area, which means that a substantial percentage of teachers do not major (or minor) in the subject area they teach. Ingersoll (1999) estimated that one-third of mathematics teachers and one-fourth of English teachers do not major or minor in the subject area they teach. Another input often seen as important is experience. On average, according to the Department of Education study, teachers have 15 years of experience.

National surveys also provide some information on the professional development teachers receive. According to the Department of Education study (National Center for Education Statistics, 1999a), nearly all teachers (99%) receive some professional development each year. The most common topics of this professional development are curriculum and performance standards (81% of teachers), educational technology (78%), new teaching methods (77%), and in-depth learning in the subject area (73%). The least common topic is dealing with limited-English-proficient and culturally different students (31%). Most professional development appears not to be very intensive, typically involving a one-day workshop. The only topic for which a majority of the teachers being trained receive more than eight hours of professional development is in-depth learning in the relevant subject area (56%).

Finally, another Department of Education study tabulated classroom practices for the nation as a whole (National Center for Education Statistics, 1999b). It described four types of classroom practice: interactions between teachers and students, materials and resources used in the classroom, the nature of the learning tasks students perform in school and at home, and methods for assessing student progress. In terms of interactions, the study found that most teachers use a combination of interactions with the class as a whole, interactions with small groups of students, and one-on-one interactions with individual students. Most teachers also use a combination of lecturing, talking with students, and having students talk among themselves. Materials and resources commonly used in classrooms consist of printed materials, concrete materials, and educational technology. Of printed materials, teachers are equally likely to use textbooks or supplementary materials in class. The use of textbooks predominates in homework, however. Two out of three teachers also make use of worksheets. The use of various concrete materials is also quite common. Three out of four teachers use physical models and other objects; nine out of 10 use blackboards or overhead projectors; and more than half use computers. In terms of learning tasks, the study found a mixed picture. In classwork, about three out of five teachers engage in activities involving higher-order thinking skills. Homework, however, only involves higher-order activities 13% of the time; two-thirds of homework activities are drills. Finally, information on assessing student progress reveals that slightly more than half of all teachers are engaged in some kind of on-going portfolio assessment (57%). Interestingly, these portfolios, which involve collections of student work over time, often include point-in-time assessments—namely, tests (62%).

While the information from these surveys does provide a sketch of classroom practices, professional development, and teacher inputs, it does not relate them to student outcomes, such as scores on standardized tests. One exception is an analysis of the National Educational Longitudinal Study of 1988 (NELS:88; National Center for Education Statistics, 1996). To produce the NELS:88 database, a national sample of students was surveyed when they were in eighth grade, then resurveyed in the 10th and 12th grades. The surveys included assessments in mathematics, science, history, and reading. The students' teachers, principals, and parents were also surveyed. The analysis related the information on classroom practices reported by mathematics and science teachers to student test scores in those two subjects for nearly 10,000 students, taking into account various background characteristics of the students and other factors that might potentially influence test scores. The study found that students performed better in mathematics when their teachers emphasized higher-order thinking skills. In science, emphasizing higher-order thinking skills did not make a difference.

While the NELS:88 analysis is thus suggestive of a link between classroom practices and student academic performance, it is hampered by various shortcomings in the database. The list of classroom practices in the database is extremely limited; for instance NELS:88 does not include information on hands-on learning or the prevalence of tests in classrooms. The database also includes very little professional development information, and as a result, professional development is not related to student test scores in the analysis. Also, like the input-output studies mentioned earlier, the analysis does not take into account the multilevel nature of the data.

The current study seeks to remedy these problems in relating classroom practices to student achievement by applying multilevel techniques to a more recent database with more comprehensive information on classroom practices, the National Assessment of Educational Progress (NAEP). Before presenting the results of the current study, however, it is worthwhile to investigate what NAEP can add to the other national surveys of the prevalence of teacher inputs, professional development, and classroom practice.

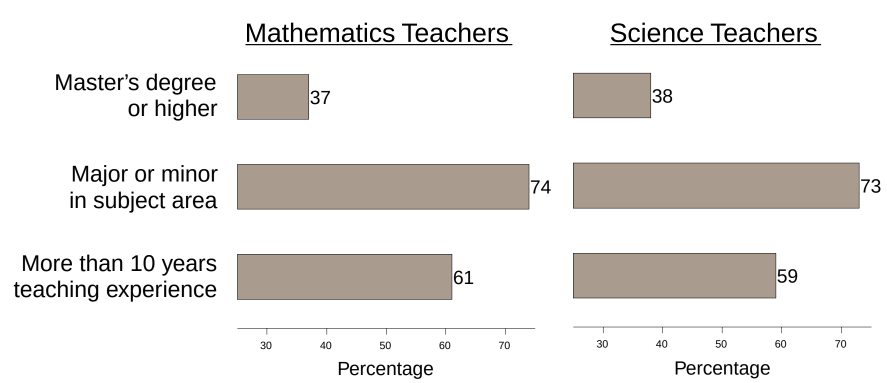
Chapter Two: A Portrait of America's Teachers and Their Classroom Practices

The data used in the current study for the description of teacher inputs, professional development, and classroom practices, as well as for linking them to student achievement in mathematics and science, come from the National Assessment of Educational Progress (NAEP). Known as "the Nation's Report Card," NAEP is administered every year or two to national samples of fourth, eighth, and twelfth graders in various subjects, including mathematics, reading, science, history, and geography. In addition to administering an assessment in the subject area to students, NAEP sends questionnaires to students, teachers, and school administrators.[1] All of this information is used to measure trends in student performance over time and to compare performance among subgroups of students, such as males and females.

For this study, the eighth graders who took the 1996 assessments in mathematics and science were selected for analysis. In total, 7,146 eighth graders took the mathematics assessment and a different group of 7,776 eighth graders took the science assessment. The mathematics and science databases included students' scores on the assessments, background information about the students (such as their socioeconomic status) drawn from the student questionnaires, and information on teacher inputs, professional development, and classroom practices, as well as other school information (such as class size) drawn from the teacher questionnaires. Because it comes from a national sample, this information makes it possible to describe teacher inputs, professional development, and classroom practices, as well as their impact on test scores in mathematics and science, for the nation as a whole. It should be kept in mind, though, that because the sample is of eighth graders, these results pertain to the middle school level only.

NAEP reveals that many eighth-grade math and science teachers are lacking in certain inputs (Figure 1). One out of three of these teachers has at least a master's degree, which means that the remaining two out of three do not.[2] Three out of four of these teachers have majors or minors in

Figure 1
Characteristics of Eighth-Grade Mathematics and Science Teachers



Source: 1996 National Assessment of Educational Progress in Mathematics and Science.

1 NAEP surveys students' teachers in the relevant subject area.

2 NAEP is nationally representative of students, but not necessarily of teachers. Consequently, the percentages reported here should not be understood to mean the percentages of teachers with a certain characteristic, but the percentage of students whose teachers have that characteristic.

their subject area (mathematics or mathematics education for mathematics teachers and science or science education for science teachers), which means that one out of four does not. Finally, six out of 10 teachers have more than 10 years of experience, which means that four out of 10 lack such experience.

NAEP shows that mathematics and science teachers receive some professional development in a variety of topics (Figure 2). NAEP asks mathematics teachers whether they have received any professional development at all in the last five years in nine topics, and asks science teachers the same question for eleven topics. For mathematics teachers, the most common form of professional development concerns cooperative learning; seven out of 10 received such training in the last five years. Half received professional development in interdisciplinary instruction. Only a minority of math teachers was exposed to the remaining topics. Just under one-half received instruction in higher-order thinking skills; four out of 10 received instruction in classroom management, performance-based assessment, and portfolio assessment; and less than one-third received professional development in topics involving responsiveness to differences in the student population, namely dealing with cultural diversity, limited English proficiency, and special needs.

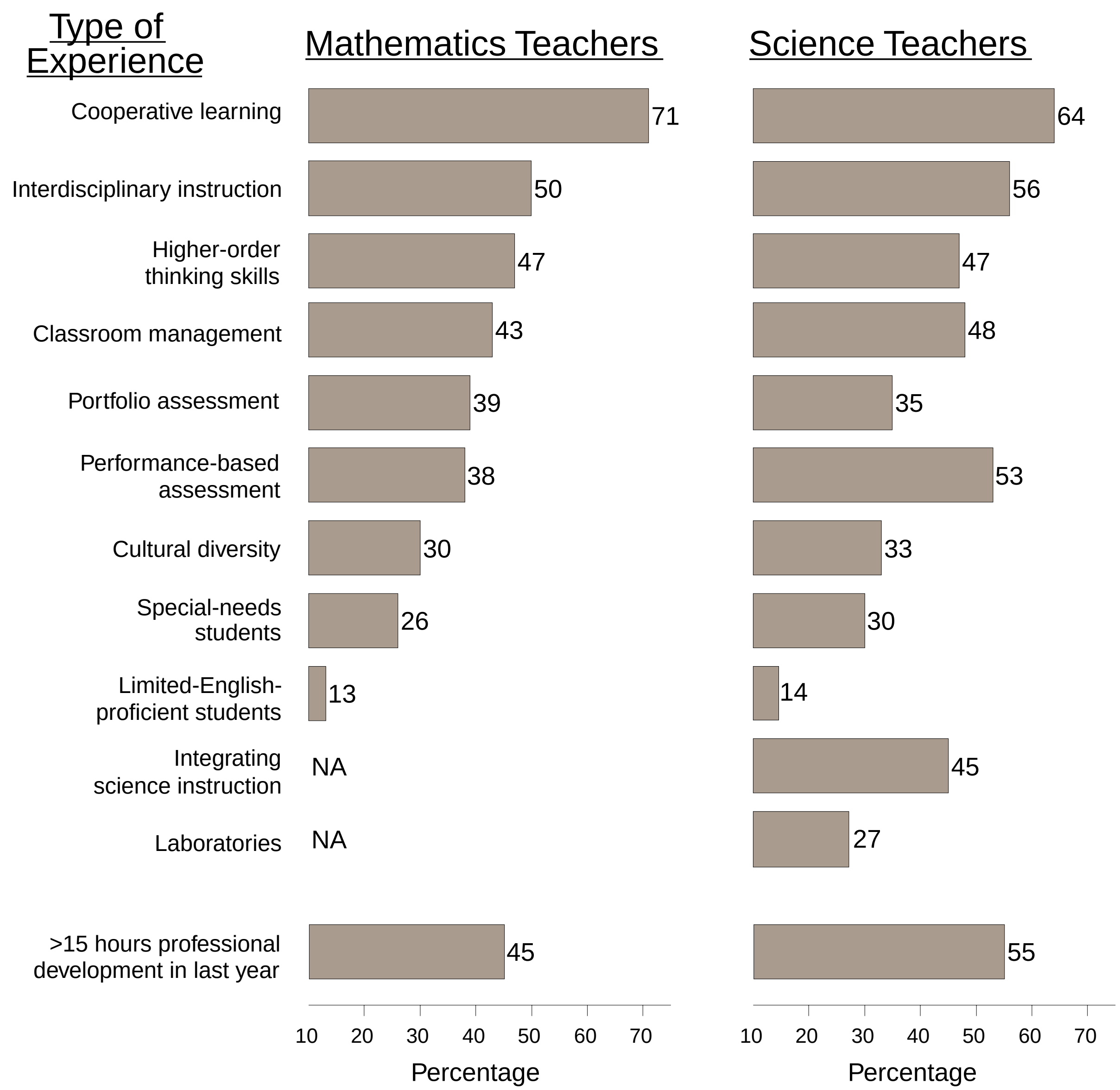
By and large, a similar pattern is revealed among science teachers. A majority of teachers received professional development in cooperative learning and interdisciplinary instruction. Less than one-half but more than one-third received professional development in higher-order thinking skills, classroom management, and portfolio assessment. And one-third or less of teachers received professional development in dealing with different student populations. One topic in which the experiences of math and science teachers differ is performance-based assessment: 38% of math teachers received professional development on this topic, as opposed to 53% of science teachers. Science teachers were also asked about other professional development topics. NAEP indicates that 27% of science teachers received professional development in laboratory skills and 45% in integrating science instruction.

These patterns indicate significant gaps in professional development among the teaching force. Because teachers are only asked whether they have received any professional development at all for each topic, a "yes" answer could indicate anything from a two-hour seminar to a semester-long course. For most topics, a majority of teachers did not even reach this minimum threshold. And for those that did, the professional development they received may indicate only a superficial acquaintance with the topic. This possibility is further supported by the results from another question asked of teachers: How much time did they spend in all forms of professional development over the last year? About one-half of mathematics and science teachers reported spending more than 15 hours in professional development. This means that half of the teachers receive no more than two days worth of professional development a year.

NAEP asked teachers about three aspects of classroom practices: the knowledge and skills they sought to convey, the activities and methods they used in class to accomplish this, and the types of assessments they used to monitor student progress. In mathematics, the most advanced knowledge covered in eighth grade tends to be algebra and geometry (Figure 3). Among math teachers, 57% report covering some algebra and 24% some geometry. Moreover, teachers are much more likely to convey lower-order skills than higher-order skills. Four out of five teachers report spending time solving a series of routine problems, whereas just one out of two reports spending time applying concepts to solve unfamiliar or unique problems.

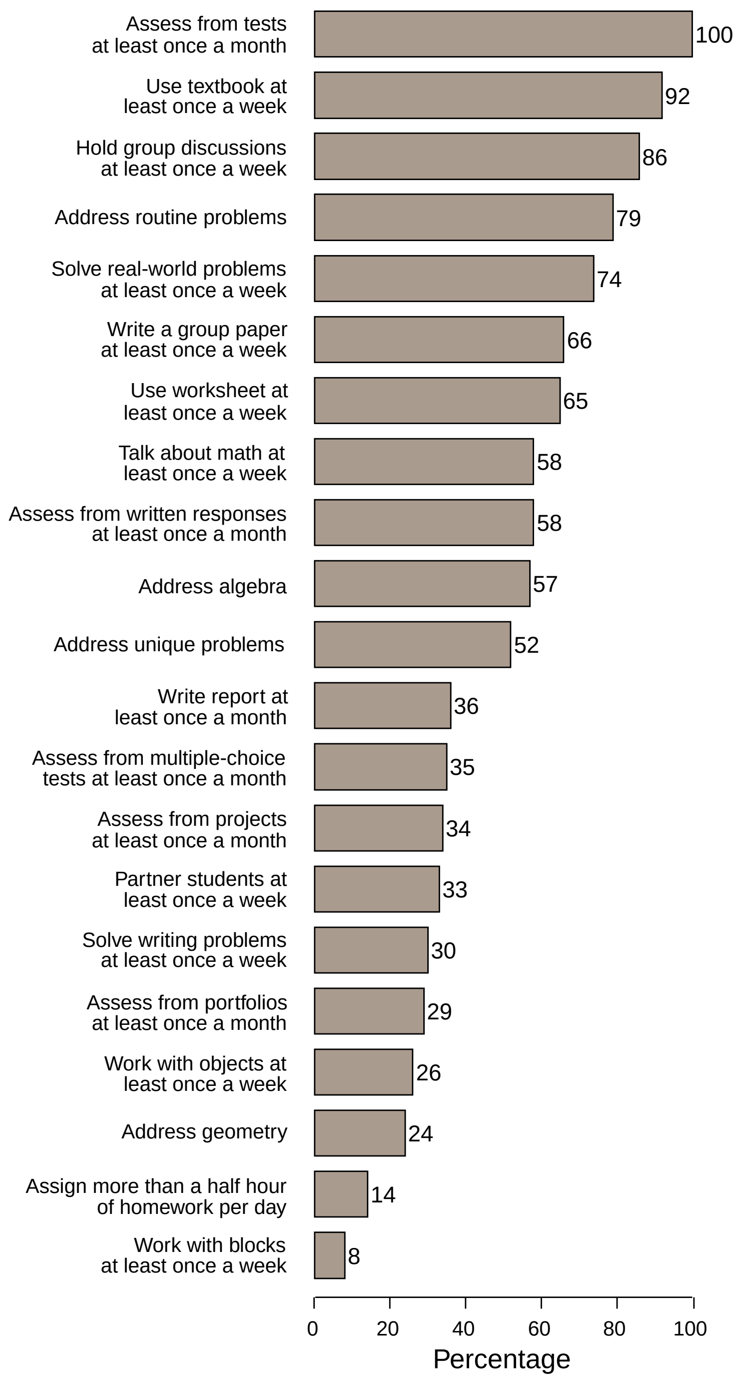
The activities and methods of mathematics teachers included assigning homework, working in groups, using written materials such as textbooks or worksheets, having students write about mathematics, solving problems involving real world situations, engaging in hands-on activities, and discussing math. Eighth graders are generally assigned between one and 30 minutes of homework a night; only 14% of their teachers report assigning

Figure 2
Professional Development Experience of Eighth-Grade Mathematics and Science Teachers in the Last Five Years



Source: 1996 National Assessment of Educational Progress in Mathematics and Science.

Figure 3
Classroom Practices of Eighth-Grade Mathematics Teachers



Source: 1996 National Assessment of Educational Progress in Mathematics.

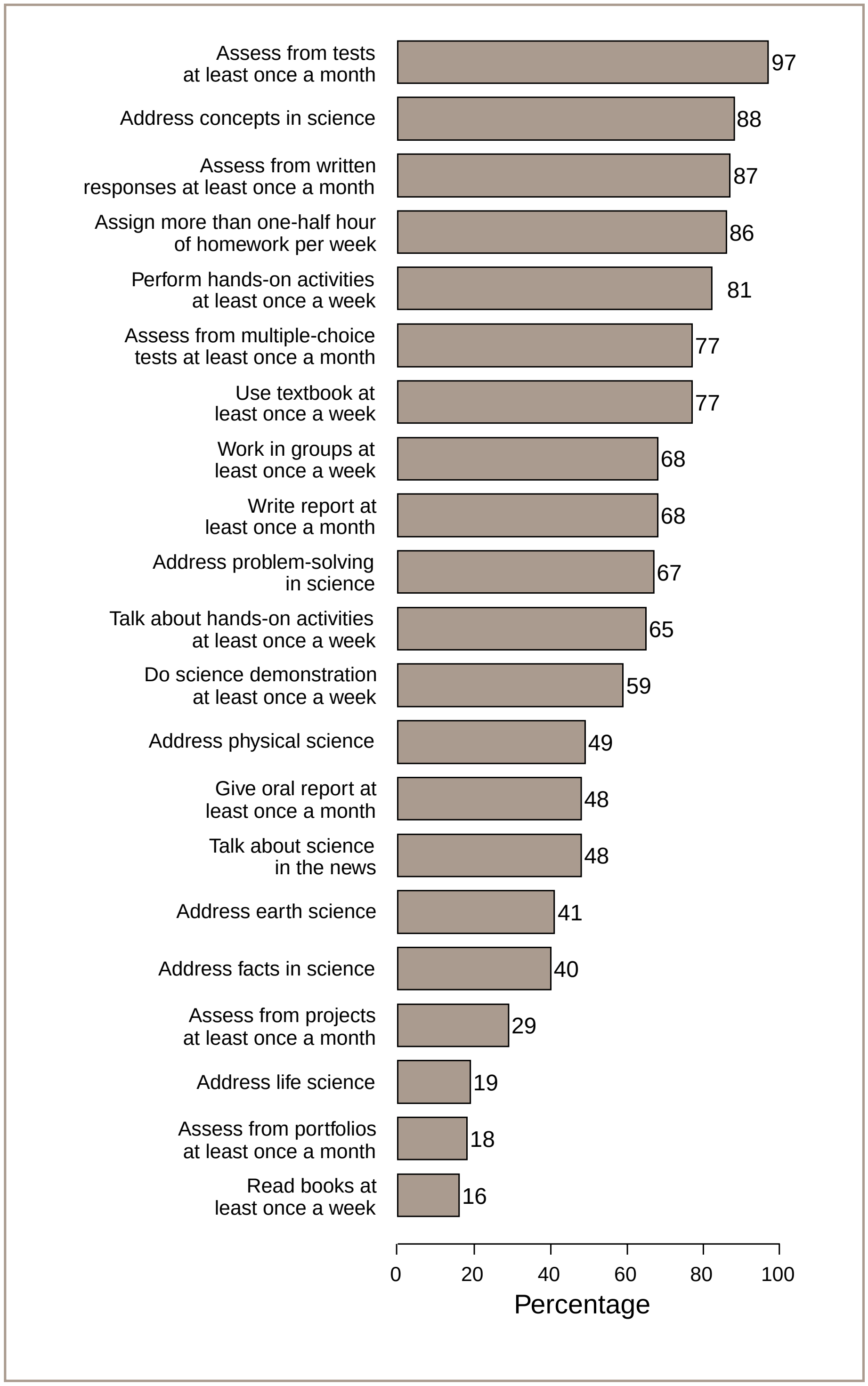
more than a half hour. Group activities vary in their frequency. Most teachers report having students discuss mathematics in groups (86%) and assigning group papers (66%). But just 33% report having students solve problems with partners. Of the written materials, textbooks are the most commonly used (92% of teachers). Two out of three teachers use worksheets. Writing about mathematics is fairly uncommon. Just 36% of teachers report assigning written reports about mathematics, and only 30% report assigning problems that involve writing. Solving real world problems is fairly common; three out of four teachers report assigning such problems at least once a week. Hands-on activities are fairly uncommon; just 8% of teachers report working with blocks and 26% with objects such as models. Class discussion of math occurs at least once a week for 58% of the teachers.

Assessments of student progress can take a variety of forms in a math class. Students can be given tests or quizzes at particular points in time. The tests thus measure retrospectively what the students have learned since the last test. Such tests can take the form of multiple-choice questions or questions requiring more extended written answers (referred to as “constructed responses”). Students can also be assessed through on-going work. One approach is to collect a portfolio of student work, from worksheets to written reports. Another approach is to assign a project which brings into play all of the skills the teacher wishes the student to acquire. Of all of these forms of assessment, tests are the most common. Nearly 100% of teachers administer some sort of test at least once a month. However, most tests do not conform to the pattern of being simply multiple-choice. Only 35% of teachers report using a multiple-choice format at least once a month. More common are tests that call for constructed responses, with 58% of teachers administering this type of assessment on at least a monthly basis. On-going assessments are the least common, with approximately one-third of teachers using these to monitor student progress.

In science, the knowledge and skills measured by NAEP are physical science, earth science, life science, conveying facts, conveying concepts, and teaching problem-solving skills (Figure 4). Of the three subjects, physical and earth science are most commonly taught (49% and 41% of teachers, respectively, report addressing these subjects). Just 19% of teachers address life science. Of the three skills, concepts are most frequently conveyed (88% of students' teachers) followed by problem solving (67%) and facts (40%).

Of the activities and methods, homework appears to be the most common, and reading books the least common. Nearly nine out of 10 teachers report assigning more than a half hour of homework each week. Students also frequently work in small groups; 68% of

Figure 4
Classroom Practices of Eighth-Grade Science Teachers



Source: 1996 National Assessment of Educational Progress in Science.

teachers report this practice. Of the written materials, textbooks are used most frequently (77% of teachers) whereas reading books or other supplementary materials occurs less frequently (16% of teachers report this). Students also often give reports—more typically written reports (68% of teachers), but also oral reports (48%). Six out of 10 teachers report demonstrating a science concept at least once a week. Hands-on activities occur with similar frequency, with 65% of teachers reporting discussing hands-on activities, and 81% reporting doing hands-on activities. About one-half of teachers report discussing science in the news with their students.

In monitoring student progress, science teachers place heavy reliance on tests at discrete time points (97% at least once a month). These tests most often involve constructed responses (87% at least once a month), but also frequently involve multiple-choice questions (77% at least once a month). Portfolios and projects are used infrequently. Three out of 10 teachers assess students using projects, and two out of 10 assess students using portfolios.

The classroom practices of mathematics and science teachers seem to differ primarily in three respects: Science teachers engage in more hands-on learning, assign more writing, and make more use of traditional forms of assessment than mathematics teachers. Among math teachers, 26% report working with objects at least once a week and only 8% report working with blocks at least once a week. By contrast, 81% of science teachers report conducting hands-on activities. Writing is relatively rare in math class, with 36% of teachers assigning written reports, whereas 68% of science teachers assign such work. And while both mathematics and science teachers seem to rely more heavily on tests than on portfolios or projects to assess student progress, math teachers are more likely to use on-going forms of assessment. Assessment from portfolios is reported by 29% of math teachers, as opposed to 18% of science teachers, and assessment from projects is reported by 34% of math teachers, as opposed to 29% of science teachers.

These descriptions provide a mixed picture of the extent to which the kinds of effective classroom practices that the qualitative literature suggests are important are actually being practiced. First, the practice of higher-order thinking skills occurs differently in mathematics and science. In mathematics, higher-order thinking skills seem to be taking a back seat to rote learning. Teachers are more likely to assign routine problems than to teach students to apply concepts to new problems. Teachers also spend little time making concepts concrete through simulations; little time is spent on hands-on activities, although real world problems are often assigned. Higher-order thinking skills seem to play a greater role in science. Science teachers are much more likely to convey concepts or problem-solving techniques than facts. It is also quite common for them to make concepts concrete through hands-on activities and demonstrations. Second, collaborative learning seems to be prevalent in both math and science; students often seem to work in groups. Third, learning seems not to be very individualized, if professional development is any guide. It is very rare for teachers to receive professional development to help them deal with the diverse backgrounds of their student bodies, be it training in issues regarding special-needs children or training in issues regarding limited-English-proficient students. Finally, assessments tend to be more of the point-in-time type than the on-going type. Thus, some effective practices are being implemented and others are not.

The foregoing discussion raises the question of whether the practices understood to be effective indeed are so. "Effectiveness" implies that these practices would improve student academic performance. Are applications, simulations, collaboration, individualization, and on-going assessment indeed associated with high student performance? This is the question the next chapter seeks to answer, by linking teacher inputs, professional development, and classroom practices to students' scores on assessments of mathematics and science.

Chapter Three: Linking Aspects of Teacher Quality to Student Test Scores

The purpose of this study is to map out the ways in which three aspects of teacher quality—teacher inputs, professional development, and classroom practices—influence student academic performance, as well as one another. This is accomplished through the statistical technique of multilevel structural equation modeling (MSEM). Like most techniques that are referred to as "multivariate," MSEM makes it possible to isolate the influence of any given factor on an outcome, taking into account the other potential influences. For instance, classroom practices can be related to student test scores, taking into account the various teacher inputs and types of professional development that might also influence these scores. In addition, MSEM makes it possible to relate a set of factors to one another, telling a story about how the outcome of interest is influenced. For instance, professional development on a given topic may not simply influence student test performance; it may also encourage teachers to engage in certain classroom practices that then translate into improved student performance. In sum, MSEM can be used to generate flow charts that indicate how various aspects of teacher quality influence one another, and how these myriad influences culminate in improved student academic performance.[3]

For this study, five sets of potential influences on student achievement are taken into account: teacher inputs, teacher professional development, classroom practices, student socioeconomic status, and class size. From NAEP it is possible to measure three teacher inputs: the number of years of teaching experience, whether the teacher has obtained a master's degree or higher, and whether the teacher majored or minored in the relevant subject area (math or math education for math and science or science education for science).

The second set of influences consists of aspects of professional development. In addition to a measure of the amount of time spent in any type of professional development over the last year, this set of influences includes six sets of professional development topics drawn from nine measures in mathematics, and eight sets of professional development topics drawn from eleven measures in science. The mathematics topics are classroom management, cooperative learning, working with different student populations (measured from professional development in cultural diversity, limited-English- proficient students, and special- needs students), on-going forms of assessment (measured from professional development on performance-based and portfolio assessments), higher-order thinking skills, and interdisciplinary instruction. The science topics include these six topics, with the addition of laboratory skills and integrating science instruction.

The third set of influences consists of classroom practices. In mathematics, a total of 12 types of practice based on the 21 measures shown in Figures 2 and 3 are included. The amount of work conducted in groups is measured by asking teachers how often students engage in group discussions, write group papers, and work on problems with partners. The extent to which written materials are used is measured by asking teachers how often students use textbooks and worksheets. The extent to which students write about math is measured by asking teachers how often students write reports and solve problems that involve writing. The extent to which students are immersed in concrete activities is measured by

3 For a more extended discussion of MSEM, see Appendix.

asking teachers how often students solve real-world problems, work with blocks, and work with other physical objects. The extent to which students take point-in-time assessments is measured by asking teachers how often students take any type of test, as well as how often they take multiple-choice tests and constructed-response tests. The extent to which students participate in on-going assessments is measured by asking teachers how often they utilize portfolios or projects for assessment purposes. Assigning homework, covering algebra and geometry, drilling students on routine problems, solving unique problems, and talking about math in class are all measured by asking teachers about each of these practices directly. In science, 14 kinds of practices based on 21 measures are included. The extent to which students use written materials is measured by asking teachers how often students read textbooks or other books. The extent to which students give reports is measured by asking teachers how often they assign written or oral reports. The extent to which students are immersed in concrete activities is measured by asking teachers how often they do demonstrations, talk about hands-on activities, and have students do hands-on activities. The extent to which students take point-in-time tests is measured by asking teachers how often students take any type of tests, as well as how often they take multiple-choice tests and constructed-response tests. The extent to which students participate in on-going assessments is measured by asking teachers how often they utilize portfolios and projects for assessment purposes. Assigning homework, working in groups, covering life, earth, and physical science topics, dealing with facts, concepts, and problem solving, and talking about science in the news are all measured by asking teachers about each of these practices directly.

The fourth and fifth sets of influences do not involve teacher quality. The fourth set—the relative affluence of students, or socioeconomic status—is measured by asking students about the education levels of their parents, as well as whether their families possess certain resources: newspapers, magazines, more than 25 books, and an encyclopedia. The fifth set—one aspect of the level of resources in the school—is measured by asking teachers how many students are in each class.

These five sets of factors are related to student academic performance in three steps. First, the impact of teacher inputs above and beyond class size and student socioeconomic status is measured. Second, professional development is added to the mix; teacher inputs, class size, and socioeconomic status are related to professional development, which, in turn, is related to student academic performance. Finally, classroom practices are included. Teacher inputs, class size, and socioeconomic status are related to professional development and classroom practices. Professional development is then related to classroom practices, which are related to student academic performance. The story of teacher quality's influence on student academic performance thus consists of three parts: Teacher inputs influence professional development, professional development influences classroom practices, and classroom practices influence student achievement. And all of these influences take into account socioeconomic status and class size, meaning that the impact of teacher quality is measured above and beyond these non-teacher-quality factors. The remainder of this chapter presents the results of these analyses, first by showing flow charts of the pathways through which aspects of teacher quality influence one another and student achievement, excluding from the charts those aspects that do not have any influence. Then the size of the influence of these aspects of teacher quality is measured using grade levels to indicate whether the influence is substantial enough to be of policy interest, and using another number to indicate the influence of each factor relative to the others.

The flow chart for mathematics indicates that one teacher input, two aspects of professional development, and two classroom practices have a positive impact on student achievement, whereas one classroom practice has a negative impact (Figure 5). Both non-teacher-quality factors influence math achievement. More affluent students score higher than less affluent students. Students in smaller classes outperform students in larger classes. Of the three teacher inputs, only the teacher's major plays a role. Students perform better when their teachers have majored or minored in the subject they are teaching. Of the six professional development topics, two prove influential. Students whose teachers receive professional development in working with different student populations outperform students whose teachers lack professional development on this topic. Also, students whose teachers receive professional development in higher-order thinking skills outperform students whose teachers lack such professional development. Of the 12 classroom practices, three have an influence on mathematics achievement. Students who frequently engage in hands-on learning seem to outperform those who spend less time in this manner. Students who are frequently exposed to higher-order thinking skills outperform those who lack such exposure. And students who participate in on-going assessment activities actually perform worse than other students.

These factors also seem to have complex interrelationships with one another. The amount of time spent in professional development in the last year is included in the chart because, although it does not directly influence mathematics achievement, it does influence a classroom practice that itself positively influences student achievement: Teachers with more professional development seem to be more likely to engage in hands-on learning activities. Socioeconomic status and class size influence factors other than student achievement. More affluent students are less likely to engage in hands-on learning activities, less likely to have teachers who have received professional development in working with different student populations, and more likely to have teachers who have spent less time in professional development overall. Students in smaller classes are also less likely to have teachers who have received professional development in working with different student populations and more likely to have teachers who have spent less time in professional development. On the other hand, teachers who have majored in the relevant subject area tend to spend more time in professional development. They are also more likely to convey higher-order thinking skills. Professional development in such skills also seems to influence classroom practice; teachers receiving such professional development are more likely to engage in hands-on learning and to use on-going forms of assessment.

In science, one teacher input, one aspect of professional development, and two classroom practices have a positive influence on science achievement, whereas one aspect of professional development has a negative influence (Figure 6). Socioeconomic status and class size again influence science achievement. More affluent students and those in smaller classes have an edge over other students. As with mathematics, the only teacher input to make a difference is the teacher's major. The students of teachers who majored in science (or science education) outperform those whose teachers had other majors. Of the eight professional development topics, the only one to positively influence science achievement is laboratory skills; students do better when their teachers have received this form of training. On the other hand, students do worse when their teachers have received professional development in classroom management.[4]

Of the 14 classroom practices, two prove important: students

[4] This negative relationship may be due to selection effects. For instance, teachers who are unskilled in managing their classes might be more likely to be steered toward this type of professional development.

Figure 5
Links Among Teacher Inputs, Professional Development, and Student Performance in Mathematics

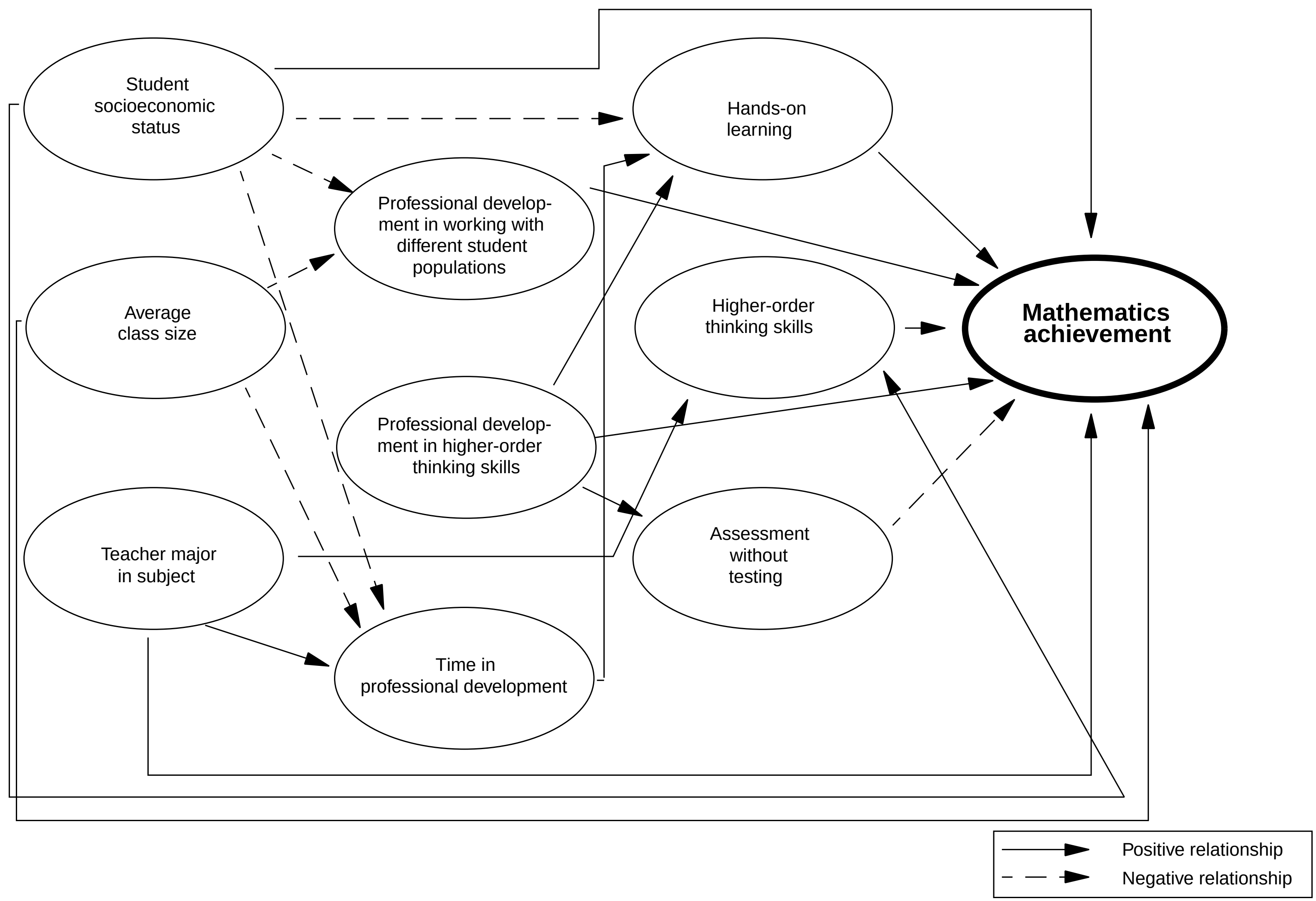
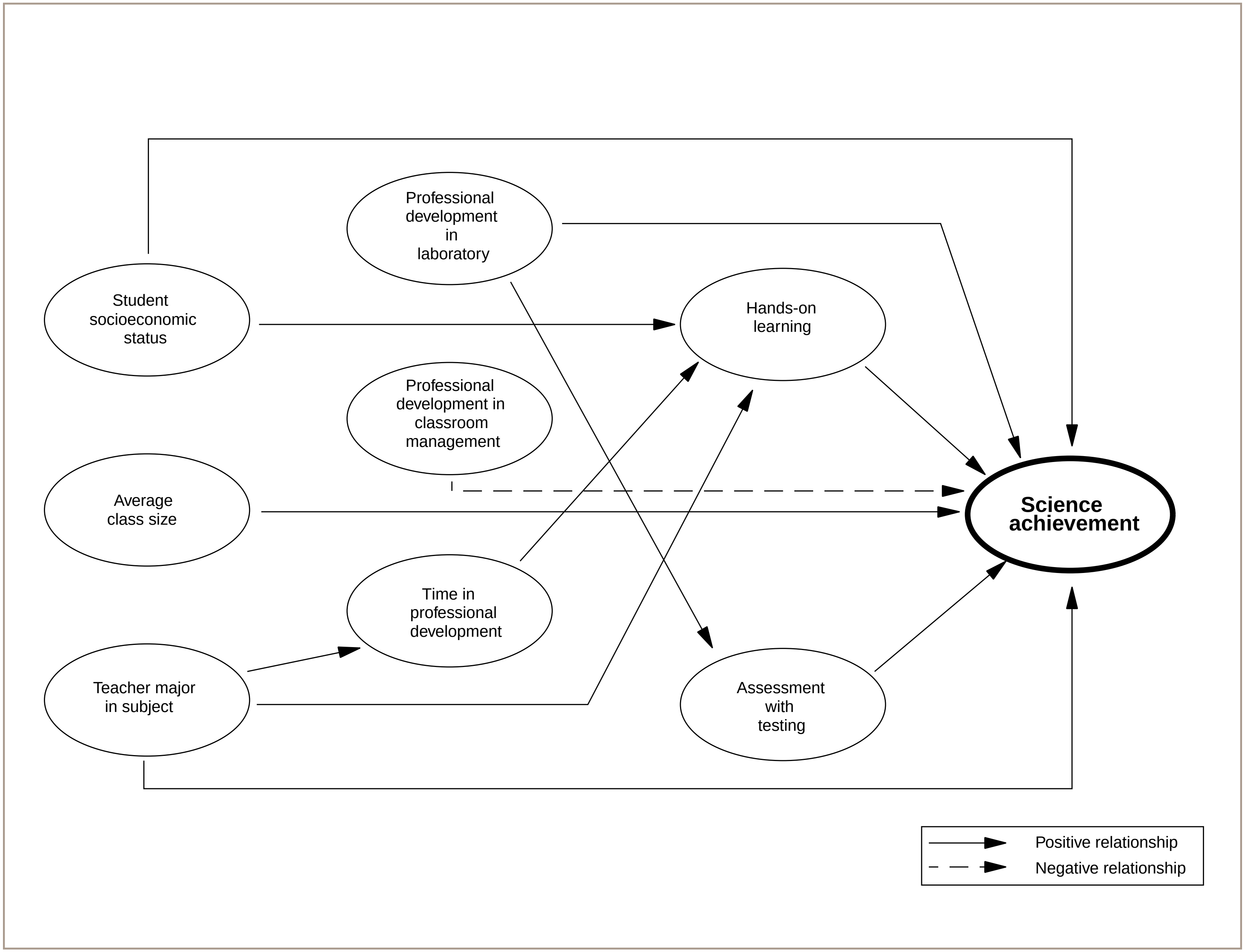


Figure 6
Links Among Teacher Inputs, Professional Development, and Student Performance in Science



whose teachers conduct hands-on learning activities and those whose teachers utilize point-in-time assessments outperform their peers. Both of these practices are thus important in both subjects; hands-on learning is important in mathematics as well, and the finding of point-in-time assessments having a positive influence in science means the same thing as the finding of on-going assessments having a negative influence in mathematics.

Some of these factors are also related to one another, although less so than in math. In science, more affluent students tend to engage in hands-on learning. Teachers majoring in the relevant subject area tend to spend more time in professional development and are more likely to engage in hands-on learning. As in math, the more time teachers spend in professional development, the more time they and their students engage in hands-on learning. And teachers who receive professional development in laboratory skills are more likely to use point-in-time assessments.

Although the flow charts indicate the important roles various teacher inputs, aspects of professional development, and classroom practices play in student learning, they do not quantify these roles. The size of the impact of each aspect of teacher quality can be measured in two ways: by the percentage of a grade level a student's test score will increase when a certain factor is present, and by the relative influence of each factor on student performance, measured using a common scale.

All of the factors that influence mathematics and science achievement seem to have an impact that can be regarded as substantial in terms of grade levels (Tables 1 and 2). Students whose teachers major in the relevant subject area are 39% of a grade level ahead of other students both in math and science. Students whose teachers receive professional development in working with different student populations are 107% of a grade level ahead of their peers in math. Students whose teachers receive professional development in higher-order thinking skills are 40% of a grade level ahead of students whose teachers lack such training in mathematics. Students whose teachers receive professional development in laboratory skills are 44% of a grade level ahead of those whose teachers lack such training in science. Students whose teachers receive professional development in classroom management are 37% of a grade level

Table 1
Impact of Teacher Inputs, Professional Development, and Classroom Practices on Mathematics Achievement: Grade Levels

Aspect of Teacher Quality	Grade Level
Major/minor in math/math education	39%
Professional development in working with different student populations	107%
Professional development in higher-order thinking skills	40%
Hands-on learning	72%
Higher-order thinking skills	39%
Assessment without testing	-46%

Table 2
Impact of Teacher Inputs, Professional Development, and Classroom Practices on Science Achievement: Grade Levels

Aspect of Teacher Quality	Grade Level
Major/minor in science/science education	39%
Professional development in laboratory skills	44%
Professional development in classroom management	-37%
Hands-on learning	40%
Assessment with testing	92%

Table 3
Impact of Teacher Inputs, Professional Development, and Classroom Practices on Mathematics Achievement: Relative Scores

Factor	Math	Science
Student socioeconomic status	.76	.75
Average class size	.10	.11
Major/minor in subject area	.09	.09
Professional development in working with different student populations	.21	.00
Professional development in higher-order thinking skills	.12	.00
Professional development in laboratory skills	N/A	.13
Professional development in classroom Management	.00	-.13
Hands-on learning	.25	.18
Higher-order thinking skills	.13	.00
Assessment without testing	-.18	.00
Assessment with testing	.00	.21

behind their peers in science. When students are exposed to hands-on learning on a weekly rather than a monthly basis, they prove to be 72% of a grade level ahead in mathematics and 40% of a grade level ahead in science. When students are exposed to higher-order thinking skills "a lot" rather than "some," they prove to be 39% of a grade level ahead in mathematics. Students exposed on a more frequent basis to on-going types of assessment in mathematics lag 46% of a grade level behind those who are exposed to these testing practices on a less frequent basis. Students exposed to point-in-time assessments in science on a more frequent basis are 92% of a grade level ahead of those exposed to point-in-time assessments on a less frequent basis.

The relative importance of various aspects of teacher quality is most clearly revealed by examining measures of their impact on test scores using a common scale (Table 3). Consistent with prior studies as far back as the Coleman Report, the data indicate that the most influential single measure is socio-economic status, with a score of .76 in math and .75 in science. However, when added together, the aspects of teacher quality (rows 3 through 11 of Table 3) are about as strong an influence, totaling .98 in math and .74 in science (negative influences are added as positive influences, meaning that they have a positive impact when the opposite phenomenon occurs). They far overshadow class size, which has a score of .10 in math and .11 in science. Of the aspects of teacher quality, classroom practices are the most important, followed by professional development, with teacher inputs being the least important. In mathematics, classroom practices (rows 8 through 11 of Table 3) have a total influence of .56 , followed by the four aspects of professional development which total .33 and teacher inputs at .09. In science, scores for classroom practices total .39, those for professional development total .26, and those for teacher inputs total .09.

In sum, the MSEM reveals the overwhelming importance of teacher quality to student academic performance. Both in mathematics and science, various aspects of teacher quality do have a substantial impact on test scores. Moreover, it appears that it was important to include classroom practices in the study, as these proved to have a larger impact than any other measures of teacher quality. Before discussing what these findings imply, however, they need to be interpreted in more detail.

Chapter Four: Conclusions

The findings of this study can be most easily interpreted by placing them in the context of prior qualitative work on classroom practices. That work noted the efficacy of teaching that conveyed higher-order thinking skills, used methods that were collaborative and individualized, and monitored student progress through on-going forms of assessment.

This study found strong support for the notion that conveying higher-order thinking skills leads to improved student performance. At first glance, the study seems to suggest that higher-order thinking skills are important in math but not science. But the measures of higher-order thinking skills used in this study involve the process of generalizing through applications. Concepts are developed from one set of problems, then applied to help solve a very different set of problems. Yet, another factor included in the study, hands-on learning, actually represents another aspect of higher-order thinking skills—namely, concretizing concepts through simulations. As hands-on learning proved to influence both math and science achievement, it would seem that the usefulness of *applications* of higher-order thinking skills is particular to math, but that the usefulness of *simulations* of higher-order thinking skills is general to both subjects. This finding is consistent with the analysis of NELS:88, which measured higher-order thinking skills primarily in terms of generalizing from one set of problems to another, and found an impact for math but not science. The NELS:88 analysis did not, however, include hands-on learning.

The current study also found some support for the effectiveness of individualizing instruction to take into account the differing knowledge and skills which different students bring into the classroom. Professional development in cultural diversity, teaching students with limited English proficiency, and teaching students with special needs were all linked to higher test scores in mathematics. In science, however, these topics of professional development did not seem to make a difference.

This study did not find support, however, for the views of the qualitative literature on collaborative learning and assessment practices, however. With regard to collaborative learning, the study revealed no benefits in either subject from working in small groups. With regard to assessment practices, the study actually found the opposite of what the qualitative research would expect. While the qualitative work supports the idea that portfolios and projects are more effective as assessment tools than point-in-time tests, this study found that students suffer academically from a lack of point-in-time testing. It would be over-reaching, however, to interpret this result as meaning that portfolios and projects should be taken out of the classroom and replaced with multiple-choice tests. The virtue of point-in-time tests is that they are the only situation in which students are on their own; their work product is to a large extent a function of their own knowledge and understanding, rather than a mixture of contributions from them, their teachers, and their peers. Being able to distinguish the work of each student from that of others has important ramifications in that it helps the teacher better understand individual students' strengths and weaknesses and makes it possible to hold students accountable for their learning. This being the case, portfolios and projects may not be adequate as tools for the assessment of an individual student because they cannot as easily distinguish the contribution of the student from that of his or her peers and teacher. But portfolios and projects may still be important in helping teachers assess what the class as a whole has learned. In this sense, portfolios and projects may be excellent tools for professional development, helping teachers identify their own strengths and weaknesses. They may also be excellent tools for teacher accountability, serving as one basis for evaluations of their work. Also, while the study found tests to be effective assessment tools, it did not find any format to be particularly effective. Multiple-choice tests are no more effective than constructed-response tests, or indeed any other kind of format. Thus, while the study supports the use of point-in-time tests, it does not necessarily support the notion that portfolios and projects should not be used, nor that multiple-choice tests should be used.[5]

Some of the findings about nonclassroom practices are also worth noting. The study finds that professional development is closely linked to classroom practice; various types of professional development encourage effective practices. In addition, the study suggests that the more extended the professional development, the more it encourages effective classroom practices. And teachers who are more knowledgeable about the subject area they teach, as measured by majoring or minoring in that subject, are also more likely to engage in effective classroom practices. All of these notions are supported by the qualitative literature, which suggests that teachers with greater mastery of their subject and armed with richer and more sustained professional development, are better able to teach higher-order thinking skills and engage in related practices, such as hands-on learning.

The methods employed to arrive at these findings did address many of the problems in prior research. The study relates a set of classroom practices to student achievement, a more comprehensive set than has previously been available. For the first time, this analysis includes professional development and teacher inputs along with these practices, showing how professional development and teacher inputs can both contribute to student performance and to effective classroom practices; it accomplishes this with a national sample, rather than the small samples that are typical in input-output research; it includes a sophisticated measure of student outcomes—scores on the standards-based NAEP tests in math and science; it takes into account characteristics of students and schools, such as student socioeconomic status; and it employs multilevel methods to handle the multiple levels of analysis in the data.

These methodological strengths of the study notwithstanding, it has several shortcomings that leave ample room for further research. First, this study covers students at only one grade level and in two subjects. Eighth graders in math and science are analyzed. Results might differ at other grade levels and in other subjects. Subsequent research should examine the math and science assessments for fourth and twelfth graders, and other subject area assessments for all three grade levels.[6]

Second, this study is cross-sectional, not longitudinal. Cross-sectional studies collect information at a single point-in-time, whereas longitudinal studies follow a group of students across many years. The disadvantage of cross-sectional studies such as this one is that the outcomes occur at the same time as the factors that apparently influence them, raising the possibility that the outcomes influence the factors rather than the other way around. Perhaps students who do well on tests tend to gravitate to classes with more hands-on learning, for example. Also, without a prior measure of student test scores, eighth-grade scores may represent the cumulative impact of school and family influences over the child's lifetime. If this is the case, then many practices of eighth-grade teachers that influence student test scores might appear not to, because they are overshadowed by the impact of the practices of earlier teachers on students' test scores. But in this case, the classroom practices of eighth-grade teachers that do appear to influence student test scores are all the more influential, because they emerge on top of the influence of the classroom practices of earlier

---

5 It should also be noted here that because these findings pertain to various kinds of tests, it is not clear whether they pertain primarily to teacher-written tests used for instructional purposes or also include state- or district-mandated tests used for accountability purposes.

6 Such a project is under way for reading using the 1998 NAEP reading assessment. Because of the noncomparability of some measures of teacher quality between the 1996 NAEP math and science assessments and the 1998 NAEP reading assessment, results from the reading assessment are not included here.

teachers. Subsequent research should address these issues by analyzing a national longitudinal sample. Since none of the existing national longitudinal databases includes a comprehensive set of classroom practices, such a study would have to collect new data. But, given the importance of classroom practices, such an effort would be well worth undertaking.[7]

Third, more detailed information on classroom practices is needed. The 1996 NAEP databases used in this study include information on whether professional development was received in various topics, but not on how much time has been spent on each topic. Also, while the databases indicate whether there is professional development on dealing with special populations of students, they do not document what teachers do in the classroom to adapt to the differing needs of their students. Perhaps subsequent administrations of NAEP could include more questions at this level of detail.

Finally, researchers need to measure the impact of classroom practices on outcomes other than tests scores. The tests used in this study are a better proxy for student understanding than many other tests would be. They include questions that tap into higher-order thinking, as indicated by the fact that test scores are related to the classroom practice of conveying of higher-order thinking skills. The NAEP tests are also designed to be consistent with national academic standards, making them useful as measures of whether students are meeting those standards. Nonetheless, it would be worthwhile to examine other important student outcomes, such as scores on Advanced Placement Tests, and outcomes that do not involve tests at all, such as how far students go in school and their occupations 10 years out of school.

These methodological shortcomings not withstanding, certain policy implications can be drawn from this study. First, the study provides ample evidence that the interest of policymakers in improving teacher quality is justified. Some policymakers maintain that there is little that school systems can do to improve student academic performance; how students perform depends too much on factors that are outside the purview of the school, such as socioeconomic status. Yet, this study indicates that one aspect of schools, the quality of their teaching force, does have a major impact on student test scores—indeed an impact that is comparable in size to that of student socioeconomic status. Other policymakers maintain that schools can make a difference, but primarily through increasing the quantity rather than the quality of teachers, thus reducing class sizes. Yet, this study indicates that the potential benefits to students of smaller class sizes, while substantial, are far overshadowed by the potential benefits of improved teacher quality. The aspects of teacher quality measured here have an impact seven to 10 times as great as that of class size.

But if the focus of policymakers on teacher quality is justified, their focus on non-classroom aspects of teacher quality is not. Policymakers seeking to improve teacher quality often propose to alter the qualifications or finances of teachers, hoping that such changes will result in improved student achievement. This study finds, however, that two of the three teacher inputs measured here proved unrelated to student academic performance. And the one input that did make a difference, whether the teacher majored or minored in the relevant subject, had only a modest effect, far overshadowed by the effects of classroom practices and professional development. Other inputs not included in this study, such as the preservice training of teachers or their proficiency in pedagogical knowledge, as measured by standardized test scores, might very well make a difference. Nonetheless, policymakers should stop relying on inputs that do not make a difference, and pay greater attention to the classroom practices that do. To improve teacher quality in ways that will improve student achievement, then, policymakers need to find ways to encourage effective classroom practices.

7 Measuring student test scores longitudinally is often referred to as the "value-added" approach.

Another implication of this study is that professional development is a useful tool for improving classroom practices. This study indicates that the most effective classroom practices involve conveying higher-order thinking skills and engaging in hands-on learning activities. The study also finds that teachers who receive rich and sustained professional development generally, and professional development geared toward higher-order thinking skills and concrete activities such as laboratories particularly, are more likely to engage in effective classroom practices. Policymakers could thus improve teacher quality by providing more opportunities for teachers to receive professional development. That professional development should occur over an extended period of time rather than being limited to a weekend seminar, and it should cover topics closely tied to classroom practices.

Policymakers might also make use of more prescriptive mechanisms to encourage the teaching of higher-order thinking skills and related practices. Policymakers could reform state math and science standards to make sure that they are consistent with the call of national math and science standards for greater emphasis on higher-order thinking skills. Also, provided that teachers receive adequate professional development, they might be provided with rewards for carrying this emphasis into the classroom. Financial bonuses could be provided to teachers, not simply for student test scores (a portion of which are influenced by factors outside teachers' control), but also for putting into practice a curriculum oriented toward effective classroom practices. Advanced certification, such as that of the National Board of Professional Teaching Standards, could also be offered as a reward for effective teaching along the lines described here.

In sum, this study shows not only that teachers matter most, but how they most matter. For too long, policymakers have sought to improve teacher quality by changing the mix of who goes into teaching, making teaching more attractive through higher salaries, requiring more education, or streamlining recruitment procedures. Yet what really matters is not where teachers come from, but what they do in the classroom. And it is possible to make improvements in classroom practices with the current teaching force, irrespective of educational levels or other qualifications. The first step is for policymakers to stop scratching the surface of teaching and learning through superficial policies that manipulate teacher inputs, and instead roll up their sleeves and dig into the nature of teaching and learning by influencing what occurs in the classroom.

Appendix: How The Study Was Conducted

This study makes use of the technique of multilevel structural equation modeling (MSEM). Its structural equation modeling aspect involves two components: measurement models that relate constructs to multiple indicators of those constructs, and path models that relate the constructs to one another. The hypothesized measurement and path models are tested against a covariance matrix of data on the observed variables, and the goodness of fit between the hypothesized models and the observed data is measured through various statistics, including goodness-of-fit indices and the root-mean-squared error of approximation. The multilevel version of a structural equation model has separate measurement and path models for student- and school-level variables, and it tests these models against a between-school covariance matrix and a within-school covariance matrix that is orthogonal to it. Effect sizes and significance tests are thus estimated in a manner that takes into account both levels of analysis (Muthen, 1994). One software package that supports MSEMs is STREAMS. STREAMS acts as a pre- and post-processor for structural equation modeling (SEM) software, such as AMOS or LISREL, loading in the elaborate syntax required for an MSEM and translating the SEM output in terms of a multilevel model (Gustafsson & Stahl, 1997).

For this study, three MSEMs were developed in each subject area using STREAMS. First, teacher inputs were related to student test performance, taking into account student socioeconomic status and class size. For both math and science, these teacher inputs were years of experience, a master's degree or higher, and a major or minor in the relevant subject area. Second, professional development activities were related to student test performance, taking into account student socioeconomic status, class size, and those teacher inputs found to be significantly related to student test scores in the first model. Student socioeconomic status, class size, and teacher inputs were also related to professional development. For both subjects, the professional development activities that were measured included cooperative learning, interdisciplinary instruction, higher-order thinking skills, classroom management, portfolio assessment, performance-based assessment, cultural diversity, special needs students, and limited-English-proficiency students, and the number of hours of professional development irrespective of topic. For science, integrating science instruction and laboratory skills were included as well. Third, classroom practices were related to student test performance, taking into account student socioeconomic status, class size, and those teacher inputs and professional development activities that were found to be significantly related to test scores. The amount of time in professional development was also included, although it had not been directly related to test scores in the second model. Student socioeconomic status, class size, teacher inputs, and professional development were also related to classroom practices. The practices included in math were those listed in Figure 3 and the practices included in science were those listed in Figure 4.

The results for these six models were combined into the path diagrams displayed in Figures 5 and 6. Relationships to mathematics achievement were drawn from the appropriate models (teacher inputs from model 1, professional development from model 2, and classroom practices from model 3). Relationships among variables were drawn from model 3. The effect sizes expressed in terms of grade level increments in Chapter Three were derived from the unstandardized coefficients from the appropriate models. The models express these coefficients as points on the NAEP scale. As there are roughly 50 points between the average fourth and eighth grader on the NAEP scale in each subject, one-fourth of this amount, or 12.25 points was treated as a grade level. The effect sizes expressed on a common scale in Chapter Three are the standardized coefficients from the appropriate models.

All models had adequate goodness-of-fit.

The current study makes use of data from the National Assessment of Educational Progress (NAEP), and some words are in order about the pitfalls in using such data. The two most serious potential pitfalls in using NAEP data involve variability in the measurement of student academic performance and variability in the sample. Measurement variability stems from the fact that only a subset of the items in the assessment is administered to each student. The student's score is imputed through a procedure known as "conditioning." Based upon information about the student and his or her school, five possible scores are developed, and are referred to as "plausible values." Any analysis of the data needs to take into account the error underlying the variability in the plausible values. Sampling variability stems from the fact that NAEP draws a stratified, clustered sample. If the data are treated as if NAEP is a simple random sample, standard errors will be underestimated, making nonsignificant differences appear significant (Johnson, 1989; Johnson, Mislevy & Thomas, 1994; Johnson, Rust & Wallace, 1994).

For this study, the solution to the first problem was to model measurement variability explicitly. All five plausible values were treated as manifest variables and a latent variable of student academic performance was generated from the manifest variables, and their corresponding error terms. The level of error estimated in this way closely matches the level of error calculated using other methods. The solution to the second problem was to estimate a design effect and inflate the standard errors accordingly.

It is worth noting that one common misconception about the analysis of NAEP data is that it is not possible to conduct analyses at the individual level. One reason that this is a misconception is that all measures except for student academic performance do not use plausible values methodology and consequently are meaningful for an individual respondent. And while the measures of student academic performance are not meaningful for individuals in absolute terms, the variance in performance among individuals can be estimated, provided that the error caused by measurement variability is taken into account.

References

Ballou, D. & Podgursky, M. (1997). *Teacher pay and teacher quality.* W. E. Upjohn Institute for Employment Research.

Bradley, A. (2000). Washington state to start testing teachers. *Education Week, 19* (28), 24, 32.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Ehrenberg, R. G. & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review, 14* (1), 1-21.

Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation, 28,* 465-491.

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7* (9).

Goldhaber, D. D. & Brewer, D. J. (1999). Teacher licensing and student achievement. In M. Kanstoroom and C. Finn (Eds.), *Better teachers, better schools,* (pp. 215-238). Washington, DC: Thomas B. Fordham Foundation.

Greenwald, R., Hedges, L. V. & Laine, R. D. (1996). The effects of school resources on student achievement. *Review of Educational Research, 57,* 415-435.

Gustafsson, J. E., & Stahl, P. A. (1997). STREAM user's guide: Version 1.7. Molndal, Sweden: Multivariate Ware.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19* (2),141-164.

Hoff, D. J. (2000). Massachusetts to put math teachers to the test. *Education Week, 19* (38),16,18.

Ingersoll, R. M. (1999). The problem of underqualified teachers in American secondary schools. *Educational Researcher, 28* (2), 26-37.

Johnson, E. (1989). Considerations and techniques for the analysis of NAEP data. *Journal of Educational Statistics, 14* (4), 303-334.

Johnson, E., Mislevy, R. J., & Thomas, N. (1994). Scaling procedures. In E. Johnson & J. E. Carlson (Eds.), *The NAEP 1992 technical report* (pp. 241-256). Princeton, NJ: Educational Testing Service.

Johnson, E., Rust, K. & Wallace, C. (1994). Weighting procedures and the estimation of sampling variance. In E. Johnson & J. E. Carlson (Eds.), *The NAEP 1992 technical report* (pp. 193-239). Princeton, NJ: Educational Testing Service.

McLaughlin, M. W. & Talbert, J. E. (1993). Introduction: New visions of teaching. In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding* (pp. 1-10). San Francisco, CA: Jossey-Bass.

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

Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22* (3), 399-420.

National Center for Education Statistics. (1996). *High school seniors' instructional experiences in science and mathematics*. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics. (1999a). *Teacher quality: A report on the preparation and qualifications of public school teachers*. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics. (1999b). *What happens in classrooms? Instructional practices in elementary and secondary schools, 1994-95*. Washington, DC: U.S. Government Printing Office.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York: Author.

Olson, L. & Hoff, D. J. (1999). Teaching tops agenda at summit. *Education Week, 19* (6), 1, 20.

Robelen, E. W. (1999). Riley outlines licensure plan for teachers. *Education Week, 18* (24), 1, 24.

Sanders, W. L. & Rivers, J. C. (1998). Cumulative and residual effects of teachers on future student academic achievement. In Education Trust, *Thinking K-16: Good teaching matters: How well qualified teachers can close the gap.*

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.

88502-005803 • RR100M10 • Printed in U.S.A.
I.N. 989495